University of Groningen

A skewed perspective on the universe

Flöss, Thomas

DOI:
10.33612/diss.1084344889

Publication date:
2024

[Link to publication in University of Groningen/UMCG research database](#)
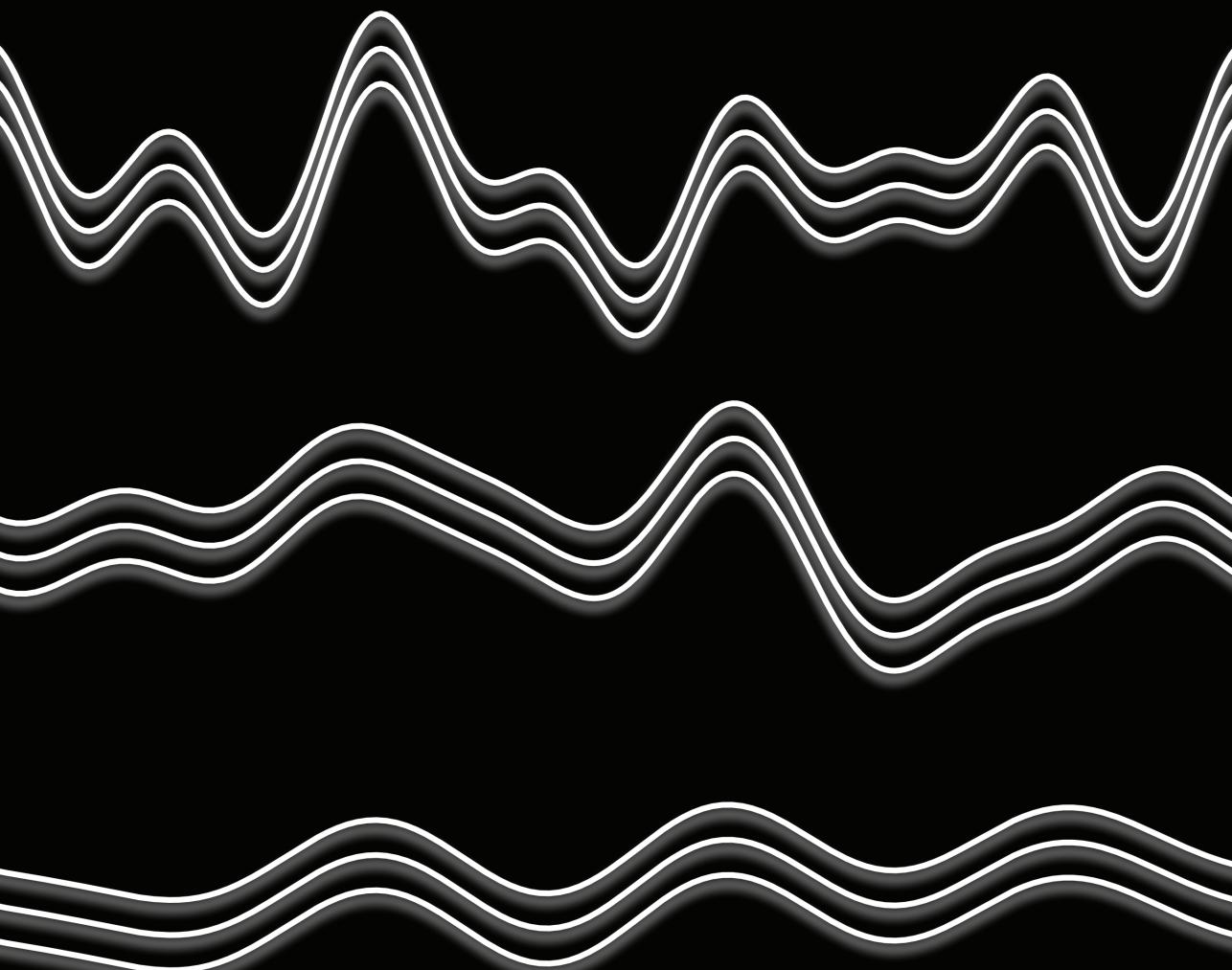
A skewed perspective on the universe
advancements, challenges and prospects in the hunt for primordial non-Gaussianity

Thomas Flöss

A skewed perspective on the universe

advancements, challenges and prospects in the hunt for primordial non-Gaussianity

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. ir. J.M.A. Scherpen,
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

dinsdag 17 september 2024 om 11:00 uur

door

Thomas Stefan Flöss

geboren op 4 december 1994

List of publications

The following papers are included in the thesis (in order of appearance in the thesis):

[1] **The Dark Ages' 21-cm trispectrum**
T. Flöss, T. de Wild, P.D. Meerburg & L.V.E. Koopmans
JCAP **06** (2022) 06, 020 [doi: 10.1088/1475-7516/2022/06/020]

[2] **Primordial non-Gaussianity and non-Gaussian covariance**
T. Flöss, M. Biagetti & P.D. Meerburg
Phys.Rev.D **107** (2023) 2, 023528 [doi: 10.1103/PhysRevD.107.023528]

[3] **Improving constraints on primordial non-Gaussianity using neural network based reconstruction**
T. Flöss & P.D. Meerburg
JCAP **02** (2024) 031 [doi: 10.1088/1475-7516/2024/02/031]

[4] **Denoising Diffusion Delensing: Reconstructing the non-Gaussian CMB lensing potential with diffusion models**
T. Flöss, W. Coulton, A. Duivenvoorden, F. Villaescusa-Navarro & B.D. Wandelt
Mon.Not.Roy.Astron.Soc. **533** (2024) 1 [doi: 10.1093/mnras/stae1818]

[5] **Non-linear electrodynamics from massive gravity**
T. Flöss, D. Roest & T. Westerdijk
JHEP **02** (2024) 194 [doi: 10.1007/JHEP02(2024)194]

Additional publications that are not part of the thesis (most recent first):

[6] **PolyBin3D: A Suite of Optimal and Efficient Power Spectrum and Bispectrum Estimators for Large-Scale Structure**
O.H.E. Philcox & T. Flöss
arXiv:2404.07249, submitted to *Phys.Rev.D*

[7] **Local non-Gaussianities from cross-correlations between the CMB and 21-cm**
G. Orlando, T. Flöss, P.D. Meerburg & J. Silk
arXiv:2307.15046, submitted to *Phys.Rev.D*

Contribution of the author to the publications in this thesis:

In [1], the author performed all the calculations and was the main writer, except for Section 2.2 and Section 2.3, the calculations and writing of which were done primarily by Tim de Wild, and have also appeared in his Master's thesis. The other authors contributed to conceptualization, methodology, writing-review, and editing.

In [2], the author performed all the calculations and was the main writer. The other authors contributed to conceptualization, methodology, writing-review, and editing.

In [3], the author performed all the calculations and was the main writer. The other authors contributed to conceptualization, methodology, writing-review, and editing.

In [4], the author performed all the calculations and was the main writer. The other authors contributed to conceptualization, methodology, writing-review, and editing.

In [5], the author performed the perturbative calculations, while all other calculations and writing were contributed to equally by all authors of the publication.

Contents

Chapter 1

Introduction

Cosmology, or the study of the universe at large, can be considered one of the greatest accomplishments of the profound developments that the field of physics underwent during the 20th century. It combines the physics of fundamental particles at the tiniest length scales, governed by the rules of Quantum Mechanics, with that of length scales as large as the entire universe, where gravity and General Relativity reign supreme. It also aims to understand both the shortest and longest timescales, from the tiniest fraction of a second after the Big Bang to the 13.7 billion years that is the age of the universe. Over the past decades, remarkable progress has been made in understanding the universe along almost its entire range of length and timescales. This section aims to provide an accessible introduction to the standard model of cosmology and some of its open questions and challenges, some of which are the topics of this thesis.

1.1   The standard model of cosmology

In 1964 Arno Penzias and Robert Wilson accidentally detected the Cosmic Microwave Background (CMB) radiation [1], a relic afterglow of the Big Bang, that can be thought of as a 'baby picture' of the universe when it was only a few hundred thousand years old. Their discovery laid the groundwork for our modern understanding of cosmology. The existence of the CMB suggests that our universe must have started in a dense, hot phase that cooled down by the expansion of space, a scenario that goes by the name Hot Big Bang. Inhomogeneities in the distribution of matter present at the time, started to grow and collapse under the influence of gravity, eventually forming stars, galaxies, and all the other structures that we observe in the universe today. Observations of these initial inhomogeneities (or anisotropies) in the CMB radiation [2], together with previous measurements of galaxy rotation curves [3, 4] and several other astrophysical phenomena, suggested that the well-known baryonic matter is outnumbered in mass about 5 to 1 by a form of matter that we are unable to observe directly, and is hence referred to as (Cold) Dark Matter (CDM). Finally, another key moment in the development of cosmology was the 1998 discovery that the universe is expanding at an accelerating rate [5, 6], suggesting the

existence of elusive vacuum energy (Λ), often referred to as a cosmological constant or Dark Energy, that makes up about 68% of the total energy content of the universe. These developments culminated in the standard model of cosmology known as Λ-CDM. Since then, ever more precise measurements of the CMB anisotropies, by WMAP [7–12], Planck [13–15], ACT [16], SPT [17], and BICEP/Keck [18] among others, have pinned down the parameters of this model with high precision. Despite this remarkable progress, several open problems remain, the most obvious of which are as to what is the nature of Dark Matter and Dark Energy. Additionally, the internal consistency of our cosmological models is in jeopardy because of disagreement between different types of astronomical observations, for example about the value of the expansion rate (or Hubble parameter). More specifically, direct distance measurements of this expansion rate, such as supernovae measurements similar to those used in the 1998 discovery [19–21], and indirect measurements, such as those of CMB anisotropies [15], conclude different expansion rates that are in tension[1].

Arguably the most fundamental question is what preceded the initial hot and dense phase suggested by the CMB, and what physical mechanism generated the initial anisotropies that allowed structure to form. In other words, **what set the initial conditions of our universe?** There are several reasons to expect that something happened before the Hot Big Bang phase and set the initial conditions. For one, the observation that the CMB temperature is a homogenous 2.73 Kelvin in every direction, with anisotropies of size around 1 part in 10000, is puzzling. Two points in the CMB that are separated by more than about one degree in our sky were causally disconnected at the time the CMB was emitted, meaning they had no means to equilibrate their temperature. From this perspective, the homogeneity of the CMB seems to be a case of severe *fine-tuning*, often referred to as the horizon problem. Another such fine-tuning lies in the fact that CMB measurements suggest that our universe today is geometrically flat. The equations of General Relativity that govern the evolution of space, tell us that for an expanding universe filled with energy, matter, and radiation, any geometrical curvature will grow over time. Hence, for the universe to appear flat today, it must have been even flatter initially, again suggesting highly fine-tuned initial conditions. Although such fine-tuning is not necessarily impossible, preferably one would like to explain these observations as a natural consequence of some more fundamental theory.

---

[1]This is commonly referred to as the Hubble tension. Another such disagreement concerns the size of matter fluctuations in the late universe, known as the $S8$ tension.

1.2 Cosmic Inflation

Around 1980, Alan Guth and Alexei Starobinsky independently sought to solve yet another puzzling fact about our universe. Theories of particle physics at the time suggested that the extreme conditions early on during the Hot Big Bang phase should have led to a significant production of magnetic monopoles [22], yet in our universe we have observed none. They proposed a brief period of exponential expansion before the onset of the Hot Big Bang phase, which could dilute any such exotic species, explaining why we never observe them. They realized that such a period of expansion would also address the aforementioned horizon and flatness problems [23, 24]. The former is solved because now our observable universe came to be in a much smaller, causally connected, and therefore homogenous patch, that rapidly expanded to cosmic scales. Furthermore, unlike the decelerating expansion that the universe undergoes after the Hot Big Bang, exponential expansion drives the universe's geometry towards flatness. This initial period of accelerated expansion is now known as cosmic inflation and has become an important pillar in the leading paradigm of cosmology. Although first formulated as a theory of false-vacuum decay, nowadays such a period of expansion is commonly thought to be realized by the early presence of one or more quantum fields, often dubbed inflaton(s), that transfer their potential energy to the spacetime in such a way as to cause exponential expansion [25, 26] [2].

Another remarkable feature of inflation is that it provides a natural mechanism for generating the tiny initial density fluctuations, that grew into the structures that fill the universe today [28, 29]. They originate from quantum fluctuations of the scalar field(s) responsible for driving the expansion. Whereas in a slowly- or non-expanding spacetime such quantum fluctuations would only live briefly before annihilating, the rapid expansion of inflation instead rips them apart over cosmic scales. The ensemble of these quantum processes taking place, results in a Gaussian (i.e. being fully described by a two-point correlation function) and nearly scale-invariant (i.e. self-similar) distribution of density fluctuations throughout the universe at the end of inflation. A visualization of this mechanism with additional explanation is given in figure 1.1. If this mechanism is indeed responsible for setting the initial conditions, then the existence of structure on cosmic scales is arguably the most macroscopic manifestation of a quantum process imaginable. Precise observations of CMB fluctuations indeed tell us that the initial conditions of the universe must have been

---

[2]The accelerated expansion of inflation is very similar to the aforementioned Dark Energy that causes our present-day universe to expand at an exponential pace. Although the energy scale, and thus the amount of expansion involved during inflation, is many orders of magnitude higher, there exist models of Dark Energy that draw inspiration from the inflationary mechanism [27].

highly Gaussian and nearly scale-invariant [30].
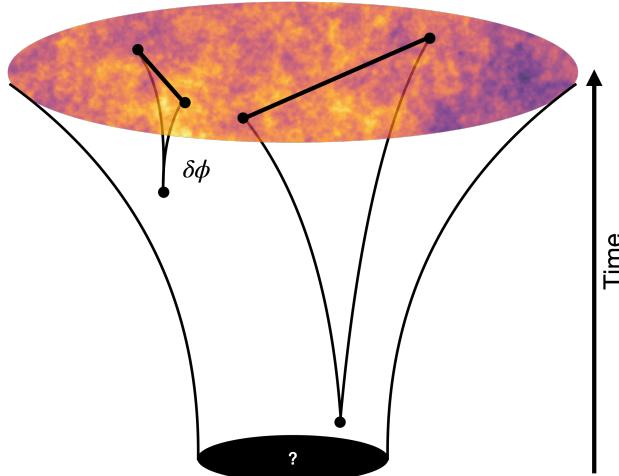
Figure 1.1: *Visual representation of how inflation generates anisotropic initial conditions. The inflaton field ϕ that drives the cosmic expansion can spawn two quantum fluctuations $\delta\phi$ that get separated to cosmic scales by the rapid expansion. The energy density at the two points of space where these fluctuations end up, become correlated. Fluctuations sourced at different times correspond to different spatial separations in the initial conditions. The ensemble of these processes induces a spatial two-point correlation, resulting in Gaussian initial conditions. Furthermore, the nearly scale-invariant nature of an exponentially expanding spacetime results in a similar (but not equal) amount of fluctuations sourced at different times, in turn resulting in similar statistical power of fluctuations at different length-scales and thus nearly scale-invariant initial conditions.*

But, despite its ability to correctly predict the statistical properties of the initial conditions, inflation is by no means a confirmed theory and more evidence is required to prove its validity. This evidence we hope to find by testing other predictions of inflation. One such prediction is the existence of *primordial tensor fluctuations*. Besides the aforementioned scalar density fluctuations, models of inflation typically generate ripples in spacetime itself, more commonly known as gravitational waves. Just as the ensemble of inflationary quantum fluctuations resulted in a distribution of scalar density fluctuations, it is expected to also produce a background of gravitational waves [23]. Although a direct detection of this background with gravitational wave interferometry (i.e. what is used to detect gravitational waves from merging black holes and neutron stars [31]) is likely out of reach for the time being [3], we hope

---

[3]At the time of writing there has been a recent detection of a gravitational wave background using

to be able to detect its imprint in the CMB where it induces a particular type of polarization of the CMB fluctuations, called B-modes, that we hope to be able to observe in addition to the already routinely measured temperature and polarization E-mode fluctuations [33].

Besides not being experimentally confirmed, inflationary models constitute a vast landscape of viable mechanisms (still) compatible with observations. Nevertheless, even in the absence of conclusive evidence, we can constrain this landscape through observational bounds on inflationary parameters. The strongest of such constraints come from bounds on a combination of the initial scalar density perturbations' deviation from scale-invariance (the *scalar spectral tilt* n_s), and the relative size of tensor perturbations to scalar perturbations (the *tensor-to-scalar ratio* r). Precise CMB measurements (including B-mode searches), combined with measurements of the distribution of galaxies in the late universe, have narrowed down the space of possible models significantly [18], as shown in figure 1.2. The next generation of surveys, Simons Observatory [34] and CMB-S4 [35], is poised to improve bounds on r by about two orders of magnitude.

1.3 Primordial non-Gaussianity and the Cosmological Collider

Another generic prediction and important probe of the physics of inflation, and the main topic of this thesis, is a possible deviation of the initial conditions from perfect Gaussianity, dubbed *primordial non-Gaussianity* [36]. Just like quantum fluctuations during inflation can give rise to Gaussian initial conditions as described previously and in figure 1.1, additional processes (i.e. interactions of fluctuations) can take place, depending on what is allowed by the inflationary theory. As visualized and explained in figure 1.3, such processes can induce three- or higher-N-point correlation functions in the initial conditions, implying a deviation from Gaussianity. These higher-N-point correlation functions are the higher-dimensional counterparts to for example the *skewness* (three-point) and *kurtosis* (four-point) of a one-dimensional probability density function (PDF). Different physical origins of such correlations will result in statistical power in distinctly *shaped* correlation functions [37], encoding valuable information about the physics at play during inflation. For example, self-interactions of the inflaton induce correlations with shapes of similar side lengths (e.g. equilateral triangles or rectangles) [38]. In contrast to this, the presence and

---

pulsar timing arrays [32]. Whether this background is due to inflation remains to be elucidated, with other options including a background of binary merger events.
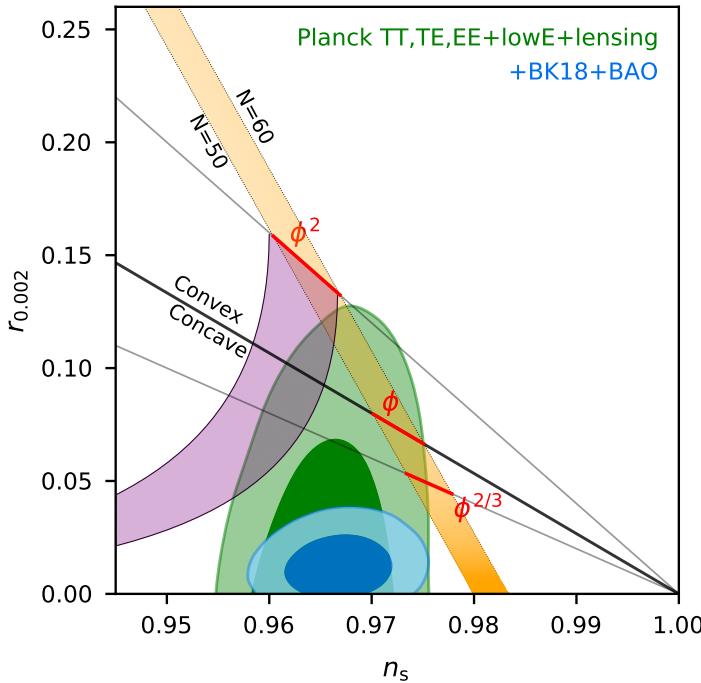
Figure 1.2: *Observational bounds on the scalar spectral tilt n_s and tensor-to-scalar ratio r from CMB (green region) and combined CMB and galaxy measurements (blue region). The other lines and regions mark different inflationary mechanisms (e.g. a quadratic potential ϕ^2) and amounts of expansion ($N = 50$ meaning 50 e-folds of inflationary expansion, i.e. physical length scales grow by 50 factors e.). Bounds on these two parameters translate into constraints on the theory space. For example, quadratic inflation (upper red line) is ruled out by these observations, since it lies well outside of the blue region. Figure taken from [18].*

interaction of multiple inflaton fields imprint correlations between short and long distances, resulting in unique shapes of non-Gaussianity, including squeezed triangle configurations known as *local* non-Gaussianity [39, 40]. A schematic visualization of this shape dependence is shown in Figure 1.4. Even in the simplest model of inflation, with a single scalar field and no self-interactions, weak non-Gaussianity will be induced by weak gravitational interaction, setting a lower limit or *gravitational floor* for primordial non-Gaussianity [41–44]. Because of its sensitivity to the details of inflation, a (non-)detection of primordial non-Gaussianity has the potential to drastically improve our understanding of inflationary physics, and extensive searches for these primordial signals are thus warranted [45–47]. For example, since multi-

field inflationary models typically predict local non-Gaussianity with a magnitude $f_{\mathrm{NL}}^{\mathrm{local}} = O(1)$, whereas we expect $f_{\mathrm{NL}}^{\mathrm{local}} \ll 1$ in single-field models, it is an important scientific target to constrain or detect the local shape at this level.
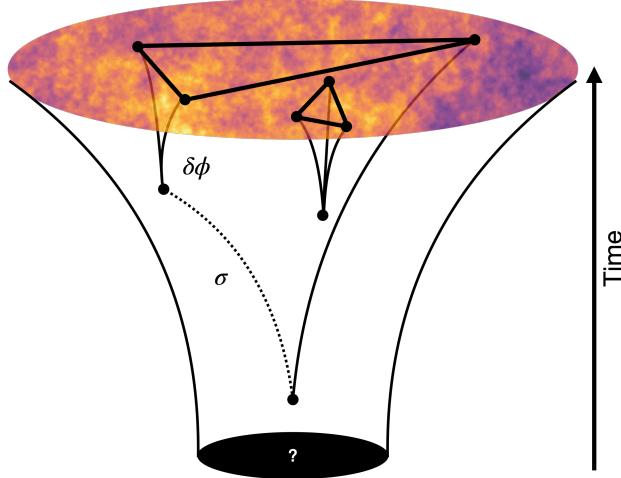
Figure 1.3: *Visual representation of the generation of primordial non-Gaussianity during the inflationary expansion. Interactions of fluctuations sourced at different times and through different mechanisms result in higher-order spatial correlation functions with different shapes. For example, long-range correlations can be mediated by an additional field σ, which can either be another light scalar (inflaton) or a more massive particle. Fluctuations sourced at the same time during inflation, result in correlations in equilateral triangles.*

More recently, it has been noted that interactions of the inflaton with additional massive fields, with masses of the order of the energy scale of inflation, can leave yet another distinct imprint in the initial conditions through primordial non-Gaussianity [48–53]. For example, through the statistical power in triangles oscillating as a function of its squeezedness, as visualized in figure 1.4. The frequency, amplitude, and phase of this oscillation are set by the mass and spin of these *spectator* fields, thereby in principle allowing for spectroscopy of the particle content of the inflationary epoch. In this sense, the inflationary expansion can be thought of as a particle collider experiment, the results of which were encoded in the initial conditions, and therefore everywhere around us in the universe. The energy scale at which inflation took place, however, is likely about eleven orders of magnitude higher than the energy scales that we can probe with the Large Hadron Collider, and conceivably out of reach for any terrestrial particle collider, ever. The *cosmological collider* therefore probes energy scales vastly beyond that of the Standard Model of particle physics, but rather closer to the Planck scale. There is indeed ample reason to expect the presence of massive

fields during inflation [54]. For one, the exponential expansion of space raises the mass of standard model particles to be close to the inflationary energy scale. Additionally, UV-complete theories of physics, such as string theory, generically predict a tower of states with masses close to the Planck scale, that could have been briefly excited during inflation, leaving their fingerprint in the initial conditions.
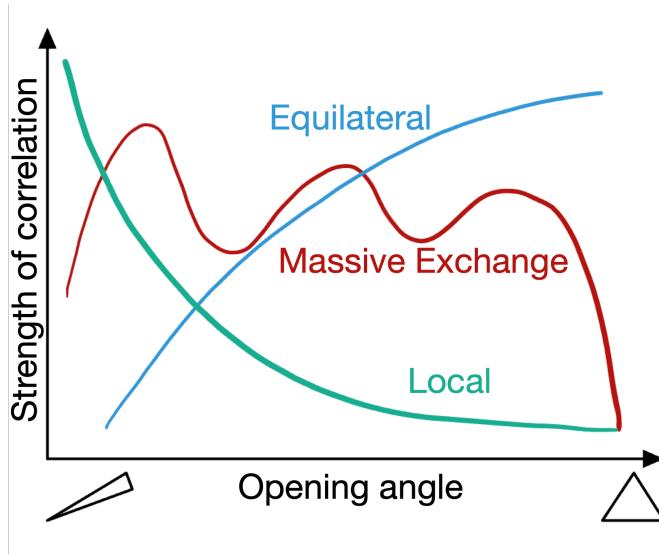
Figure 1.4: *Schematic visualization of the strength of correlation between three points in the initial conditions, for different types of primordial non-Gaussian signatures as a function of the shape of the triangle that these three points form, from equilateral triangles on the right to squeezed triangles on the left. In green, we show the long-range correlation that peaks for squeezed triangles and is known as local primordial non-Gaussianity. In blue we show the equilateral type that is typical for self-interactions of the inflaton. In red we show the oscillatory behavior of primordial non-Gaussianity induced by the exchange of a massive state as a function of the squeezedness of the triangle.*

It is thus clear that primordial non-Gaussianity constitutes a powerful window into the early universe, involving physics at energy scales that have yet to be explored. This thesis aims to contribute to the development of this intriguing and promising field of cosmology and studies advancements, challenges, and prospects in the hunt for primordial non-Gaussianity. Before we get there, in the next section we explain how we can study the initial conditions set by inflation.

1.4 Tracing the matter distribution

Unfortunately, since we are unable to look further into the past the CMB, i.e. some 300.000 years after the supposed end of inflation, we do not have direct access to even a glimpse of the initial conditions set by inflation [4]. But since the initial conditions (or more concretely the initial matter distribution) are the seeds of all structure throughout the cosmos, we can hope to learn about its properties, including primordial non-Gaussianity, by studying the observable matter distribution, or tracers thereof, across space and time. In this section, we will go over several such tracers that we can use as a probe for the initial conditions, and discuss their pros and cons.

1.4.1 The Cosmic Microwave Background

The CMB is an excellent probe for studying the initial conditions since at the time it was emitted, the underlying matter distribution has not been strongly affected by gravity, and therefore the CMB temperature fluctuations are linearly related to the initial conditions, making the CMB a pristine tracer of the initial conditions and therefore an indispensable source of information on the primordial universe. Since the CMB anisotropies are observed to be highly Gaussian, all information can be captured using the two-point correlation function, or *power spectrum* in momentum or harmonic space. The power spectrum of CMB anisotropies matches the generic prediction of inflation that the distribution of primordial fluctuations must be nearly scale invariant. Any additional higher-order correlation function, such as a three-point correlation function or *bispectrum* in momentum space, is therefore an immediate probe of primordial non-Gaussianity [55]. The relative simplicity of this analysis has resulted in a precise determination of cosmological parameters, for example using data from the Planck satellite [15]. Although so far no conclusive evidence for primordial non-Gaussianity has been found, surveys have put increasingly strong constraints on its magnitude [56], the strongest of which is $f_{\mathrm{NL}}^{\mathrm{local}} = -0.9 \pm 5.1$. Additionally, the CMB data has been searched for cosmological collider signals as well [57], again without detection.

By mapping the anisotropies of the CMB to increasing precision, the next generation of CMB experiments aims to improve sensitivity to primordial non -Gaussianity by a factor of a few, possibly closing in on a detection. However, it could well be that the amplitude of any primordial non-Gaussianity is still several orders of magnitude below the sensitivity of these experiments, down to the aforementioned gravi-

---

[4]Note that by initial conditions here we explicitly mean scalar density fluctuations. Other possible relics of the early universe, such as primordial black holes or primordial gravitational waves, could in principle be observed directly.

tational floor, and any imprint due to massive states in the cosmological collider will be very faint. Furthermore, it might not be sensitive enough to constrain the local shape at the coveted threshold of one. Naively one might hope to simply keep increasing the resolution of CMB measurements to improve sensitivity to primordial non-Gaussianity. Unfortunately, the information content of the CMB is intrinsically limited by its nature, as on small enough scales the photons that make up the CMB have been able to equilibrate the temperature across the region, washing out any anisotropies, an effect known as *diffusion damping* (or Silk damping) [58]. Hence, below a certain scale, there are simply no anisotropic features to observe in the CMB, and the next generation of CMB experiments already approaches this scale[5]. It is thus clear that to increase our chances of detecting primordial non-Gaussianity and accessing the cosmological collider, we need additional ways to probe the initial conditions.

1.4.2   Large-scale structure of the Universe

We hope to obtain further information on the initial conditions by studying the matter distribution across the entirety of its history and 3-dimensional volume. Compared to the CMB image that only constitutes the 2-dimensional boundary of this volume, its bulk contains orders of magnitude more information. As illustrated in figure 1.5, this volume covers different historical epochs of the universe as we look further and further out into space. Right after the CMB is emitted, the universe is permeated with neutral hydrogen (and some other light elements), and there are no sources of radiation that allow us to directly observe this era, hence this era is referred to as the Dark Ages. As gravity collapses the initially overdense regions of neutral hydrogen, after about 100 million years these regions become dense enough to form the first galaxies, and with it the first luminous objects (i.e. stars), ushering in a period known as Cosmic Dawn. These luminous objects emit ultraviolet (UV) radiation, heating and even ionizing the remaining neutral hydrogen gas in the interstellar medium. As more and more luminous objects form, even the gas in the intergalactic medium starts to become ionized during a period called the Epoch of Reionization. Meanwhile, matter continues to collapse and form structures for the remainder of the cosmic time until now, resulting in the Large-scale structure of the universe or Cosmic Web as visualized in figure 1.6.

For most of the cosmic volume, from today until the onset of Cosmic Dawn, there

---

[5]Let us emphasize that this does not imply that we cannot learn more from further CMB experiments. As mentioned earlier, we still hope to detect primordial tensor perturbations through CMB B-mode polarization. Even in the context of primordial non-Gaussianity, possible spectral distortions of the CMB's perfect blackbody spectrum can probe the initial conditions at completely different length scales [59].
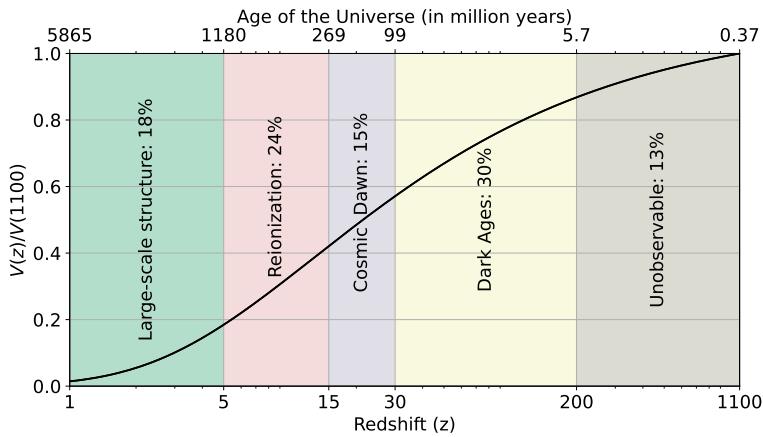
Figure 1.5: *The (cumulative) comoving volume from now (redshift z = 0) until redshift z, relative to the total volume of the observable universe (up to the surface of last scattering at redshift 1100). Different epochs of the universe are indicated, together with the percentage of the total volume that this epoch spans. The top horizontal axis shows the age of the universe at the corresponding redshifts, in millions of years. This plot aims to give a schematic visualization of the history of the universe and the approximate volume spanned by different epochs, as the precise times of the onset and end of these epochs are disputed.*
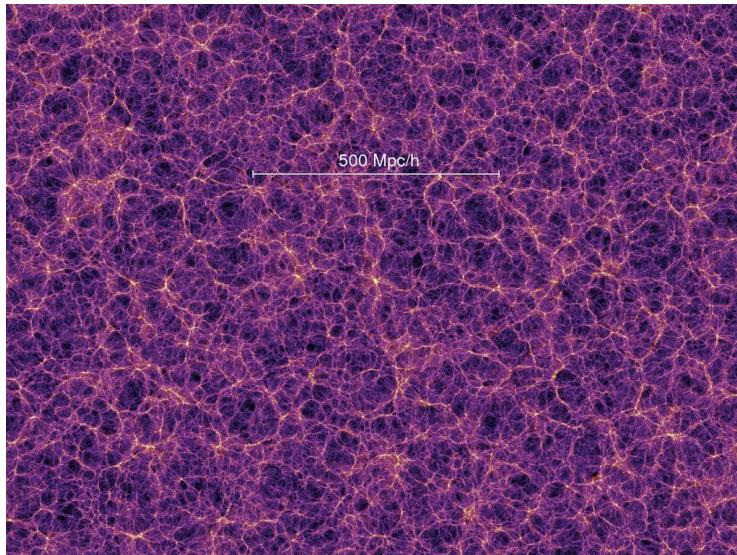
Figure 1.6: *A slice out of a high-resolution simulation of the large-scale structure of the universe, clearly showing the filaments that make up the Cosmic Web. Source: Millenium Simulations by Volker Springel.*

is a clear way for us to trace the distribution of matter: by mapping the position of luminous galaxies across different redshifts, using telescopes in a multitude of frequency ranges. Indeed this is the way to go at low redshifts, where ample such structures exist. The Sloan Digital Sky Survey's Baryon Oscillation Spectroscopic Survey (BOSS) mapped approximately 1.5 million luminous red galaxies throughout parts of the universe, as shown in figure 1.7, and many efforts are underway to expand this number by many orders of magnitude by looking deeper into space and further back in time [60–64].

1
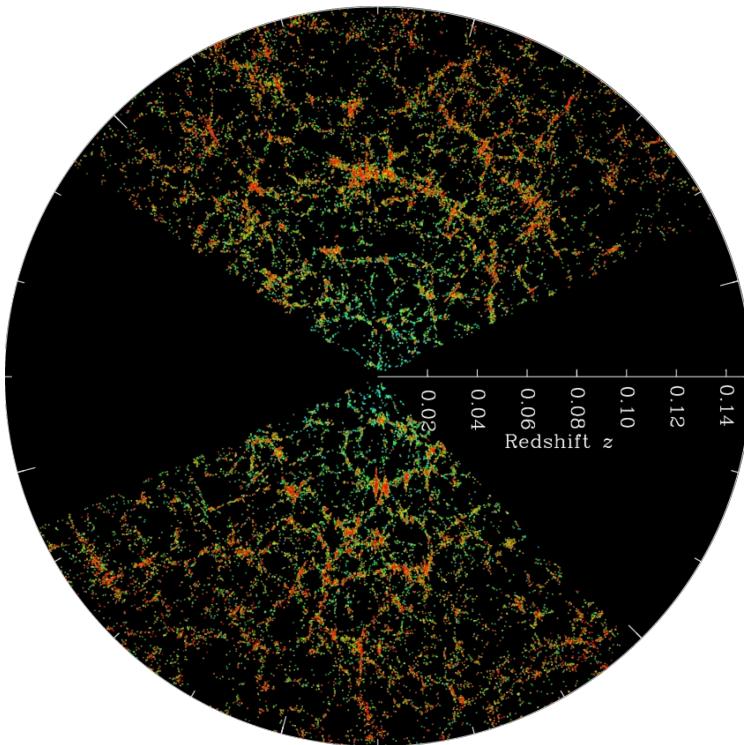
Figure 1.7: *A 2D slice through the 3D map of galaxies by Sloan Digital Sky Survey's Baryon Oscillation Spectroscopic Survey, clearly revealing Cosmic Web structures. Source: SDSS.*

Compared to the CMB, at low redshifts where galaxies can be abundantly observed, it is significantly more challenging to extract valuable information on cosmology, including the initial conditions. The universe at this age has become highly non-linear except on the very largest scales, making it very hard to make theoretical predictions to compare against the data. At redshifts close to $z = 0$, on scales larger

than about 60 Mpc (or 2×10^8 light years), we can rely on the perturbative, analytical machinery of the Effective Field Theory of Large-scale Structure (EFT of LSS) [65]. On smaller scales, matter overdensities are no longer perturbative and we have to rely on fully numerical simulations of the evolution of the universe, that also incorporate the highly non-linear formation process of bound structures, such as galaxies and clusters thereof (e.g [66]) [6]. Such simulations are computationally expensive, making them unsuitable for the Bayesian statistical methods (e.g. Markov Chain Monte Carlo) that are used to determine the best-fit parameters of our cosmological models since this requires many such simulation runs.

Additionally, the non-linear evolution of the matter density field has resulted in a highly non-Gaussian distribution of matter at late times. Unlike for the CMB, all the information encoded in the data can no longer be captured by just the power spectrum (in the absence of any primordial non-Gaussianity, that is) but instead requires also the inclusion of higher-order summary statistics, such as the bispectrum. Both measuring and modeling these higher-order correlation functions once again become increasingly complicated and computationally complex as one goes up in order and down to smaller scales. Nevertheless, it has been shown that strong ΛCDM parameter constraints can be obtained from galaxy samples such as BOSS by making use of only the power spectrum and bispectrum [67–71]. When it comes to primordial non-Gaussianity, the additional gravitational contribution to higher-order correlations also acts as confusion to any primordial signal, swamping it by many orders of magnitude, so that we are now looking for a needle in a haystack. Constraints on primordial non-Gaussianity from galaxy samples are therefore highly sensitive to uncertainties in the modeling of these *secondary* contributions to higher-order correlations. In part because of this, constraints on primordial non-Gaussianity from current generation data such as BOSS have been significantly less constraining than those from the CMB [72–75]. The next generation of galaxy surveys will yield higher-density galaxy samples, promising exquisite constraints on cosmological parameters, including primordial non-Gaussianity [64, 76, 77].

To counter the effects of non-linearity, it seems tempting to extend galaxy samples out to higher redshift, where the universe is more linear at smaller scales and thus easier to model. However, as we go to higher redshifts, these objects will become increasingly rare (as most of them have not yet formed) and hard to observe,

---

[6]This is in some sense analogous to our inability to perturbatively study strongly coupled processes in the theory of quarks and gluons, Quantum Chromodynamics (QCD), instead requiring numerical lattice simulations. Contrary to N-body simulations, for QCD it is the low-energy regime that cannot be perturbatively modeled due to quark confinement, whereas the high-energy regime is perturbative due to asymptotic freedom.

so there are fewer observations to accurately resolve the underlying matter density distribution, to the point where there are none at all, during the Dark Ages. Since even the most futuristic survey only aims to map the distribution of galaxies out to a redshift of about $z \sim 5$ [78], there is still more than 80% of the local universe's total volume left to observe, that is also significantly more linear. But with a lack of galaxies to trace the matter distribution, how to proceed?

1.4.3 The 21-centimeter spectral signature

Luckily, the neutral hydrogen that permeates the universe early on has a way of conveying its presence to us. The electron of a neutral hydrogen atom can be either aligned or anti-aligned with the spin of the nucleus. The energy gap between these states, which is known as the hyperfine transition, is about 5.86 μeV, corresponding to a photon wavelength of 21 centimeters. This hyperfine state can be excited by collisions of neutral hydrogen atoms, or the absorption of radiation, and upon relaxation emits a 21-cm photon. Depending on the temperature of the hydrogen gas relative to the temperature of the CMB radiation, an excess of these absorption or emission processes results in a 21-cm spectral line that can be observed against the black-body CMB radiation[7]. Additionally, regions of higher matter density contain more hydrogen, so that more emission or absorption processes take place, resulting in a brighter 21-cm spectral signature than that from lower-density regions. Hence, 21-cm brightness fluctuations trace the underlying matter density distribution.

During the first 6 million years of the universe, the temperature of the gas is tightly coupled to the CMB temperature so there is no 21-cm signal, making this period truly unobservable. As the universe expands further, the gas cools faster than the CMB radiation so that the spectral line becomes visible in absorption to the CMB spectrum. From this point onwards the spectral signal can be observed up until the moment the neutral hydrogen gas becomes completely ionized at the end of Reionization, thus not only making it a powerful probe of cosmology but also of the physics of Cosmic Dawn and Reionization itself [79]. Figure 1.8 shows a theoretical prediction for the brightness temperature of the volume-averaged (or global) 21-cm signal across these epochs. Many ongoing and upcoming efforts aim to measure the global 21-cm signal or its fluctuations at various redshifts [80–85], with a recent claim of detection of the signal around redshift $z = 17$ by the EDGES experiment [86].

---

[7]Technically, whether the 21-cm signal can be observed in absorption or emission depends not on the gas temperature but on the *spin temperature*, that quantifies the relative amount of excited versus non-excited neutral hydrogen atoms. If the spin temperature is higher (lower) than the CMB temperature, the 21-cm spectral line can be observed in emission (absorption) against the CMB. If the spin temperature equals the CMB temperature there is no observable spectral line.

At later times, the only remaining neutral hydrogen resides inside galaxies where it continues to emit its spectral signature, thus providing an additional probe of the distribution and formation process of galaxies [87, 88]. On this front, there have been multiple claims of detection of 21-cm fluctuations in cross-correlation with galaxy data [89–93], and one of its auto-correlation [94].
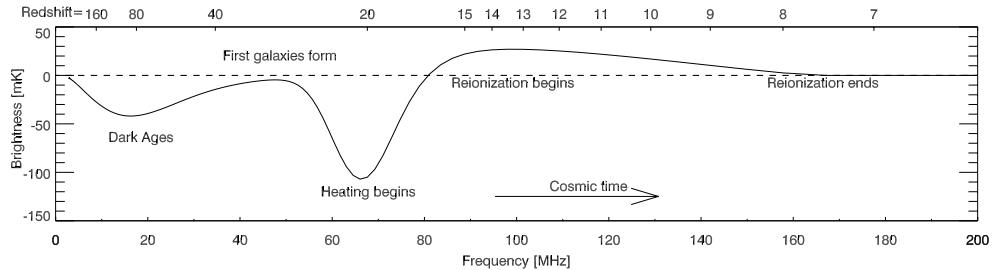
Figure 1.8: *Theoretical prediction for the brightness of the 21-cm spectral line in contrast to the CMB temperature as a function of redshift (top horizontal axis). Positive (negative) brightness temperature implies that the signal can be observed in emission (absorption) to the CMB spectrum. The bottom horizontal axis shows the frequency that the 21-cm signal has today, as a consequence of the redshift due to cosmic expansion as it travels towards us. Note that the time direction is opposite to that in figure 1.5. Source: [95]*

1.5   A light in the darkness

During the Dark Ages, the 21-cm spectral line can be observed in absorption to the CMB spectrum, in principle giving us access to an additional 30% of the cosmic volume. Most importantly, it provides a window into the matter distribution when it was still highly linear while also being clean of complicated astrophysics, making it relatively easy to model with analytical methods. This suggests the Dark Ages as the ultimate frontier for primordial non-Gaussianity, containing orders of magnitude more information than the CMB does [96]. The true challenge with doing 21-cm cosmology during the Dark Ages is not theoretical, however, but observational. First, the 21-cm signal from this era has since redshifted into the meter wavelength range, to which earth's ionosphere is opaque. We would thus have to measure it from space, most preferably from the far side of the moon, where radio interference from Earth is also minimized [97]. Secondly, the global 21-cm signal (i.e. the mean of the signal across the sky) from the Dark Ages is expected to be very faint, and thus hard to observe. Additionally, since fluctuations in the matter density field are still small, the brightness anisotropies that we hope to observe to trace the matter distribution are even weaker. To make matters worse, the signal is additionally covered

in radiation from different foreground sources, both galactic and extragalactic. Nevertheless, its promise to provide valuable insights into cosmology is tempting, and extensive investigation into its true power is warranted. Over the past two decades this has resulted in several studies aiming to provide accurate modeling of the 21-cm signal and its spatial correlation functions [98], which were then used to forecast the constraining power of the signal [99, 100], also in the context of primordial non-Gaussianity [101–103].

In Chapter 2, which is based on [104], we further investigate and develop the 21-cm signal from the Dark Ages as a probe of primordial non-Gaussianity. We establish the most accurate model of Dark Ages' 21-cm observables to date, including for the first time its four-point function or *trispectrum*. Subsequently, we use this to investigate the constraining power of the signal as a probe for primordial non-Gaussianity, providing more accurate forecasts for constraints on primordial bispectra, including cosmological collider signals. Additionally, we present the first forecasts for constraints on primordial trispectra, including collider signals. The results obtained in that Chapter further establish the Dark Ages as the ultimate probe for primordial non-Gaussianity.

1.6   A non-Gaussian hurdle

As non-linear evolution couples modes of different wavelengths, summary statistics (e.g. powerspectrum and bispectrum) of the density field become correlated. This implies that the data that we use to fit our cosmological models to, contains a reduced amount of unique information, thus lowering the precision of cosmological parameter constraints. This correlation between data is usually captured by a covariance matrix, that accounts for the overlap in information content, and, just like the summary statistics themselves, needs to be theoretically modeled. If the correlation between data is neglected, the data are assumed to be independent Gaussian random variables, whose distribution is parametrized by the entries of a diagonal covariance matrix. Any correlation between data will be captured by off-diagonal entries of the matrix, and are made up of higher-order (non-Gaussian) correlation functions, hence referred to as *non-Gaussian covariance*. Failing to properly account for data covariance can result in overconfident parameter constraints, and thus possibly false detection of new physics, such as primordial non-Gaussianity.

Covariance matrices are especially hard to compute for higher-order correlation function data. Furthermore, at higher redshifts where the density field (or a tracer thereof) is assumed to be linear to smaller scales, it is typically assumed that non-

Gaussian contributions to the covariance are small. In the context of primordial non-Gaussianity, it has therefore been common practice to neglect non-Gaussian covariance in forecasts of the constraining power of cosmological surveys, in particular, those at high redshifts, such as in Chapter 2. The impact of non-Gaussian covariance of the bispectrum at low redshifts has been appreciated for some time (e.g [105–107]), but only recently it was shown that non-Gaussian contributions to the covariance of the bispectrum are especially large in squeezed triangle configurations [108]. This therefore significantly affects constraints on local primordial non-Gaussianity, whose signal initially resides mostly in squeezed triangles. As we go to higher redshifts, such as the Dark Ages, we expect to be able to model increasingly smaller (linear) scales and access its cosmological information. This also implies that we will have access to increasingly squeezed triangles, which promises to yield strong constraints on local primordial non-Gaussianity [102, 103]. However, an investigation is warranted into the importance of non-Gaussian covariance at these redshifts.

In Chapter 3, which is based on [109], we embark on this mission. We will show that contrary to common belief and assumption, non-Gaussian covariance can significantly impact constraints on local primordial non-Gaussianity even at high redshifts. We demonstrate the consequences of neglecting non-Gaussian covariance in forecasts, and as a realistic example, we revise existing forecasts for the PUMA line-intensity mapping experiment [110, 111]. Our results show that forecasts on primordial non-Gaussianity that do not include non-Gaussian covariance, including those presented in Chapter 2 and elsewhere in the literature, need to be interpreted with caution. Hence, any effort to study local primordial non-Gaussianity using the highly redshifted 21-cm signal from the Dark Ages will additionally require new methods for more optimal analysis. One such method is described and investigated in the subsequent chapter.

1.7   Reconstructing primordial non-Gaussianity

The issue of non-Gaussian covariance is an issue of optimality of summary statistics for capturing certain information. For example, the increased covariance of squeezed triangle configuration tells us that whatever information starts in squeezed triangles (e.g. local primordial non-Gaussianity), will no longer reside there after significant non-linear evolution. Instead, the information has been moved into higher-order correlations and smaller scales. Extracting the information using spatial correlation functions therefore becomes an increasingly difficult and computationally expensive task. Thus, new techniques and tools ought to be developed to optimize the extraction of information from non-linear, non-Gaussian data, both in the context of Chap-

ter 3 and more generally. Recent work has shown that alternative summary statistics can more efficiently capture certain information encoded in the density field, e.g. [112–114].

In Chapter 4, which is based on [115], we investigate the possibility of improving the extraction of information by first reconstructing the linear (near-Gaussian) density field, starting from the late-time matter distribution at the field level. Once the field has been linearized (and Gaussianized) we expect that lower-order correlation functions (i.e. powerspectrum and bispectrum) contain an increased amount of information. To this end, we train a neural network on the output density field and initial conditions of a large set of numerical simulations [116]. To demonstrate that this indeed results in an improved information content of these spatial correlation functions, we show using a Fisher forecast that the resulting reconstruction can be used to constrain cosmological parameters, including primordial non-Gaussianity, with significantly increased accuracy, depending on the parameter.

1.8   Delensing the CMB with diffusion models

So far we have described the CMB as a pristine tracer of the initial conditions. In reality, however, the observed CMB is affected by the non-linear evolution of the matter distribution as well. As CMB photons travel towards us, their paths are deflected by the intervening matter, an effect known as weak gravitational lensing [117–121]. This results in a distortion of the observed CMB temperature and polarization fluctuations as compared to the true primary CMB and as a consequence, it is more complicated to infer primordial information from the CMB directly. Most notably, lensing induces strong secondary B-mode polarization, obfuscating the sought-after primary B-modes that are due to primordial gravitational waves [122]. Additionally, much like the non-linear evolution of the matter distribution that we have discussed previously, lensing gives rise to higher-order correlations in CMB data, affecting the extraction of primordial non-Gaussianity through both signal confusion (i.e. secondary non-Gaussianity) and non-Gaussian covariance [123, 124]. Therefore, to improve sensitivity to the primordial component, analogous to the reconstruction of the linear matter distribution discussed in Chapter 4, one can attempt to reconstruct the true CMB by *delensing* observed CMB data, thereby improving constraints on the tensor-to-scalar ratio as well as primordial non-Gaussianity [125].

In Chapter 5, we apply score-based generative models, more colloquially known as diffusion models, to the task of reconstructing the CMB lensing potential. These powerful machine-learning models can learn high-dimensional probability distribu-

tions, and subsequently sample from them, which has made them particularly powerful in the context of image generation, resulting most prominently in generative models such as OpenAI's DALL-E. In the context of CMB lensing reconstruction, these models enable probabilistic sampling of the Bayesian posterior probability distribution, similar to the state-of-the-art approach of Ref. [126]. Since our method learns the posterior entirely from the training data, we do not require an explicit expression for it. This is unlike the approach of Ref. [126], which is practically limited to the case of Gaussian lensing potentials. The additional freedom allows us to for the first time perform accurate CMB lensing reconstruction of more realistic non-Gaussian lensing potentials, that are obtained by ray-tracing N-body simulations [127]. This reconstruction provides an alternative approach to determining the CMB lensing bispectrum, which is hard to do using traditional estimators (e.g. the quadratic estimator), due to large, complicated noise biases [128].

1.9   A new connection for the Double Copy

The physical processes taking place during inflation, that give rise to primordial fluctuations and non-Gaussianity, can be seen as the cosmological counterpart to the scattering processes studied in particle accelerators. This analogy is most striking in the context of the cosmological collider. Although research on *cosmological correlators* is relatively young, the study of *scattering amplitudes* goes back to the development of Quantum Field Theory (QFT). Although such scattering amplitudes can often be computed with a tedious calculation, starting from an explicit theory (i.e. Lagrangian), the result of such a calculation can turn out to be remarkably simple. During the 1960s this observation led to an approach toward amplitudes that aimed to bypass brute-force computation, by relying on the fundamental principles of quantum field theories, such as causality, unitarity, analyticity, locality, and symmetry [129, 130]. This approach, known as the *bootstrap* program, indeed allows one to correctly determine scattering amplitudes without the need for lengthy calculations or even an explicit theory. Additionally, it provides important insight into the general properties of amplitudes [131].

Given the conceptual analogy between scattering amplitudes and cosmological correlators, it is not surprising that there exist analogies of such general properties too. A notable example is the existence of *soft theorems* due to spontaneous symmetry breaking in both scattering amplitudes [132] and cosmological correlators (e.g. [133–135]). Moreover, recent years have seen increasing efforts to study the bootstrap of cosmological correlators, which has resulted in a host of new insights into the properties of these correlators and the development of new analytical tools (e.g.

[136–139]) as well as generic predictions for primordial non-Gaussianity [140, 141]. Thus, from an observational perspective, the cosmological bootstrap program improves our understanding of the possible inflationary imprints to look out for.

Since the initial bootstrapping efforts, the scattering amplitudes community has revealed intriguing connections between fundamental theories appearing in nature. Such a connection was first observed in string theory [142], but turns out to be especially intriguing in the context of field theory. Most notably, it turns out that the scattering amplitudes appearing in the field theory formulation of gravity (i.e. General Relativity) can be constructed from the amplitudes of the theory underlying the weak and strong interactions, Yang-Mills theory [143]. Such constructions have become known as the scattering amplitude *double copy* because in this sense gravity amplitudes can be thought of as two copies of Yang-Mills amplitudes [8]. The development of the double copy program has since yielded important insights into aspects of quantum field theory and quantum gravity [144] and has revealed an entire web of related theories [145, 146]. Given the power of the double copy to simplify the construction of amplitudes, it is also of considerable interest to investigate its existence and use in constructing cosmological correlators, although much remains to be understood [147, 148].

Although not directly applicable to the cosmological context of the rest of this thesis, In Chapter 6 we demonstrate, in the context of scattering amplitudes, that two theories naturally appearing in the web of double copy theories are additionally related. More generally, we show that the low-energy effective action of linear electrodynamics (Maxwell theory) coupled to massive gravity (i.e. integrating out the massive graviton), is that of all duality invariant theories of non-linear electrodynamics. In the case of the only non-pathological (ghost-free) theory of massive gravity (due to de Rham, Gabadadze and Tolley, i.e. dRGT theory [149]), this includes the special non-linear completion of electrodynamics by Born and Infeld [150]. As visualized in figure 1.9, both Maxwell theory coupled to gravity and Born-Infeld (BI) theory appear in the web of double copy theories. From this perspective, our procedure connects these theories (red arrow). This suggests a possible deeper connection between integrating out massive degrees of freedom, and the double copy.

---

[8]The particular connection between General Relativity (GR) and Yang-Mills (YM) theory amplitudes has become known under the name *color-kinematics duality* and the punchline equation $\text{GR} = \text{YM}^2$

1
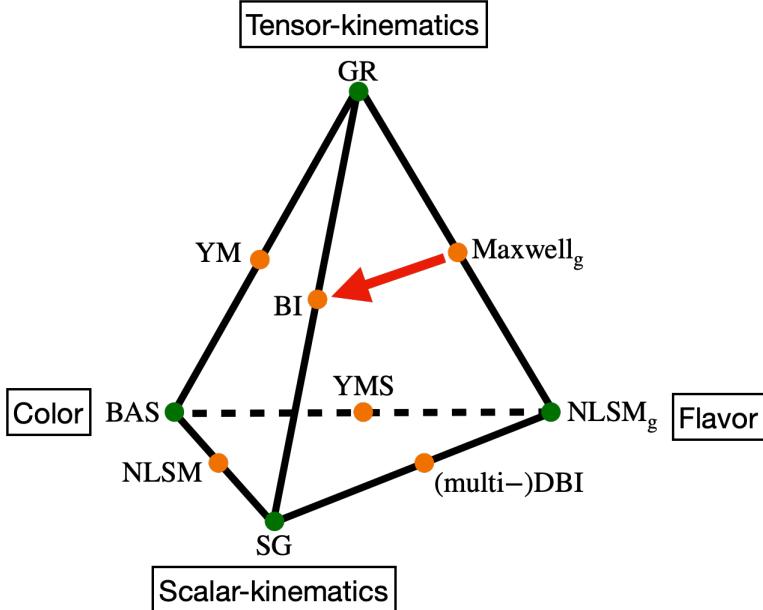
Figure 1.9: *Tetrahedron of theories related by double copy relations. Each theory is constructed out of a particular combination of four different factors (tensor-kinematics, scalar-kinematics, color, and flavor). Theories on the vertices (green) are made up of a pair of the same factors, whereas theories on the edges (orange) are made out of a combination of those at their corresponding vertices. In Chapter 6 we demonstrate an additional connection between Maxwell theory coupled to gravity (Maxwell$_g$) and Born-Infeld (BI), as visualized by the red arrow. The other theories appearing in the web are General Relativity (GR), non-linear sigma model (NLSM), Special Galileon (SG), Yang-Mills (YM), Yang-Mills Scalar (YMS), Dirac-Born-Infeld (DBI) and Bi-adjoint Scalar (BAS). Figure adapted from [146].*

1.10 Definitions, notation and conventions

Before proceeding, let us define the most important physical quantities, notation, and conventions, that will be used throughout the rest of this thesis.

Most of the thesis will be concerned with the study of fields. The inflaton field is denoted by ϕ, whereas its fluctuations are given by $\delta\phi$. To study primordial scalar perturbations we will use the comoving curvature perturbations ζ. The cold dark matter (CDM) density field is denoted by δ_c, but shortened to simply δ when baryons are neglected and cold dark matter is the only type of matter. The velocity field of cold dark matter fluctuations is additionally given as δ_v. The baryonic density field is denoted explicitly as δ_b. Any other fields, particularly in the context of 21-cm cosmology in Chapter 2, will be defined when necessary.

Fields can either be represented in real space coordinates $\mathbf{x}$, i.e. $\delta(\mathbf{x})$, or in Fourier space using wavevectors/momenta $\mathbf{k}$ as $\delta(\mathbf{k})$, which for brevity is more often denoted as $\delta_{\mathbf{k}}$. Momentum integrals such as the one appearing in the Fourier transform, are written in the following condensed notation for readability and brevity:

$$\int_{\mathbf{k}_i} = \int \frac{d^3\mathbf{k}_i}{(2\pi)^3} \tag{1.1}$$

We will often be interested in n-point spatial correlation functions of fields in the Fourier domain, where they are called spectra, e.g. powerspectrum, bispectrum, etc. Such correlation functions are denoted using angled brackets, e.g. the powerspectrum of primordial fluctuations ζ: $\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \rangle$. Due to spatial isotropy and homogeneity, spectra will only be non-zero when the momenta of fields add up to zero. To compactify our notation we therefore introduce primed (′) correlators, that exclude the Dirac delta function for the momenta and conventional prefactor:

$$\langle \zeta_{\mathbf{k}_1} \cdots \zeta_{\mathbf{k}_n} \rangle = (2\pi)^3 \delta_D(\mathbf{k}_1 + \cdots + \mathbf{k}_n) \langle \zeta_{\mathbf{k}_1} \cdots \zeta_{\mathbf{k}_n} \rangle' \tag{1.2}$$

Furthermore, the most commonly used primed correlators used in this work, that of the powerspectrum (P), bispectrum (B), and trispectrum (T), are further shortened to:

$$\begin{aligned}
\langle \zeta_{\mathbf{k}_1} \zeta_{-\mathbf{k}_1} \rangle' &= P_\zeta(\mathbf{k}_1), \\
\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle' &= B_\zeta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3), \\
\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \zeta_{\mathbf{k}_4} \rangle' &= T_\zeta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4)
\end{aligned} \tag{1.3}$$

The magnitude of a vector is denoted as $k \equiv |\mathbf{k}|$. Hats denote unit vectors $\hat{\mathbf{k}} \equiv \mathbf{k}/k$. We denote $k_1 + k_j \equiv k_{ij}$ and $\mathbf{k}_i + \mathbf{k}_j \equiv \mathbf{k}_{ij}$. Note that $k_{ij} \neq |\mathbf{k}_{ij}|$. We also define Mandelstam-like variables $\mathbf{s} = \mathbf{k}_1 + \mathbf{k}_2$, $\mathbf{t} = \mathbf{k}_1 + \mathbf{k}_4$ and $\mathbf{u} = \mathbf{k}_1 + \mathbf{k}_3$.

Bibliography

[1] A. A. Penzias and R. W. Wilson, "A Measurement of excess antenna temperature at 4080-Mc/s," *Astrophys. J.* **142**, pp. 419–421, 1965.

[2] G. Smoot *et al.*, "COBE Differential Microwave Radiometers: Instrument Design and Implementation," *ApJ* **360**, p. 685, Sept. 1990.

[3] F. Zwicky, "On the Masses of Nebulae and of Clusters of Nebulae," *Astrophys. J.* **86**, pp. 217–246, 1937.

[4] V. C. Rubin and W. K. Ford, Jr., "Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions," *Astrophys. J.* **159**, pp. 379–403, 1970.

[5] A. G. Riess *et al.*, "Observational evidence from supernovae for an accelerating universe and a cosmological constant," *Astron. J.* **116**, pp. 1009–1038, 1998.

[6] S. Perlmutter *et al.*, "Measurements of Ω and Λ from 42 High Redshift Supernovae," *Astrophys. J.* **517**, pp. 565–586, 1999.

[7] C. L. Bennett *et al.*, "The Microwave Anisotropy Probe (MAP) mission," *Astrophys. J.* **583**, pp. 1–23, 2003.

[8] D. N. Spergel *et al.*, "First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Determination of cosmological parameters," *Astrophys. J. Suppl.* **148**, pp. 175–194, 2003.

[9] D. N. Spergel *et al.*, "Wilkinson Microwave Anisotropy Probe (WMAP) three year results: implications for cosmology," *Astrophys. J. Suppl.* **170**, p. 377, 2007.

[10] E. Komatsu *et al.*, "Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation," *Astrophys. J. Suppl.* **180**, pp. 330–376, 2009.

[11] E. Komatsu *et al.*, "Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation," *Astrophys. J. Suppl.* **192**, p. 18, 2011.

[12] G. Hinshaw *et al.*, "Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results," *Astrophys. J. Suppl.* **208**, p. 19, 2013.

[13] Planck Collaboration, Ade, P. A. R., *et al.*, "Planck early results. i. the planck mission," *A&A* **536**, p. A1, 2011.

[14] P. A. R. Ade *et al.*, "Planck 2015 results. XIII. Cosmological parameters," *Astron. Astrophys.* **594**, p. A13, 2016.

[15] N. Aghanim *et al.*, "Planck 2018 results. VI. Cosmological parameters," *Astron. Astrophys.* **641**, p. A6, 2020. [Erratum: Astron.Astrophys. 652, C4 (2021)].

[16] M. S. Madhavacheril *et al.*, "The Atacama Cosmology Telescope: DR6 Gravitational Lensing Map and Cosmological Parameters," *Astrophys. J.* **962**(2), p. 113, 2024.

[17] L. Balkenhol *et al.*, "Measurement of the CMB temperature power spectrum and constraints on cosmology from the SPT-3G 2018 TT, TE, and EE dataset," *Phys. Rev. D* **108**(2), p. 023510, 2023.

[18] P. A. R. Ade *et al.*, "Improved Constraints on Primordial Gravitational Waves using Planck, WMAP, and BICEP/Keck Observations through the 2018 Observing Season," *Phys. Rev. Lett.* **127**(15), p. 151301, 2021.

[19] A. G. Riess *et al.*, "A 2.4% Determination of the Local Value of the Hubble Constant," *Astrophys. J.* **826**(1), p. 56, 2016.

[20] A. G. Riess, S. Casertano, W. Yuan, L. M. Macri, and D. Scolnic, "Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond ΛCDM," *Astrophys. J.* **876**(1), p. 85, 2019.

[21] A. G. Riess, "The Expansion of the Universe is Faster than Expected," *Nature Rev. Phys.* **2**(1), pp. 10–12, 2019.

[22] G. 't Hooft, "Magnetic Monopoles in Unified Gauge Theories," *Nucl. Phys. B* **79**, pp. 276–284, 1974.

[23] A. A. Starobinsky, "Spectrum of relict gravitational radiation and the early state of the universe," *JETP Lett.* **30**, pp. 682–685, 1979.

[24] A. H. Guth, "The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems," *Phys. Rev. D* **23**, pp. 347–356, 1981.

[25] A. D. Linde, "A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems," *Phys. Lett. B* **108**, pp. 389–393, 1982.

[26] A. Albrecht and P. J. Steinhardt, "Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking," *Phys. Rev. Lett.* **48**, pp. 1220–1223, 1982.

[27] E. J. Copeland, M. Sami, and S. Tsujikawa, "Dynamics of dark energy," *Int. J. Mod. Phys. D* **15**, pp. 1753–1936, 2006.

[28] A. H. Guth and S. Y. Pi, "Fluctuations in the New Inflationary Universe," *Phys. Rev. Lett.* **49**, pp. 1110–1113, 1982.

[29] J. M. Bardeen, P. J. Steinhardt, and M. S. Turner, "Spontaneous Creation of Almost Scale - Free Density Perturbations in an Inflationary Universe," *Phys. Rev. D* **28**, p. 679, 1983.

[30] Y. Akrami *et al.*, "Planck 2018 results. X. Constraints on inflation," *Astron. Astrophys.* **641**, p. A10, 2020.

[31] B. P. Abbott *et al.*, "Observation of Gravitational Waves from a Binary Black Hole Merger," *Phys. Rev. Lett.* **116**(6), p. 061102, 2016.

[32] G. Agazie *et al.*, "The NANOGrav 15 yr Data Set: Evidence for a Gravitational-wave Background," *Astrophys. J. Lett.* **951**(1), p. L8, 2023.

[33] U. Seljak and M. Zaldarriaga, "Signature of gravity waves in polarization of the microwave background," *Phys. Rev. Lett.* **78**, pp. 2054–2057, 1997.

[34] P. Ade *et al.*, "The Simons Observatory: Science goals and forecasts," *JCAP* **02**, p. 056, 2019.

[35] K. N. Abazajian *et al.*, "CMB-S4 Science Book, First Edition," 10 2016.

[36] T. J. Allen, B. Grinstein, and M. B. Wise, "Nongaussian Density Perturbations in Inflationary Cosmologies," *Phys. Lett. B* **197**, pp. 66–70, 1987.

[37] D. Babich, P. Creminelli, and M. Zaldarriaga, "The Shape of non-Gaussianities," *JCAP* **08**, p. 009, 2004.

[38] P. Creminelli, "On non-Gaussianities in single-field inflation," *JCAP* **10**, p. 003, 2003.

[39] D. Wands, "Multiple field inflation," *Lect. Notes Phys.* **738**, pp. 275–304, 2008.

[40] D. Langlois, "Non-Gaussianities from isocurvature modes," *J. Phys. Conf. Ser.* **405**, p. 012003, 2012.

[41] A. Gangui, F. Lucchin, S. Matarrese, and S. Mollerach, "The Three point correlation function of the cosmic microwave background in inflationary models," *Astrophys. J.* **430**, pp. 447–457, 1994.

[42] V. Acquaviva, N. Bartolo, S. Matarrese, and A. Riotto, "Second order cosmological perturbations from inflation," *Nucl. Phys. B* **667**, pp. 119–148, 2003.

[43] J. M. Maldacena, "Non-Gaussian features of primordial fluctuations in single field inflationary models," *JHEP* **05**, p. 013, 2003.

[44] G. Cabass, E. Pajer, and F. Schmidt, "How Gaussian can our Universe be?," *JCAP* **01**, p. 003, 2017.

[45] E. Komatsu *et al.*, "Non-Gaussianity as a Probe of the Physics of the Primordial Universe and the Astrophysics of the Low Redshift Universe," 2 2009.

[46] P. D. Meerburg *et al.*, "Primordial Non-Gaussianity," *Bull. Am. Astron. Soc.* **51**(3), p. 107, 2019.

[47] A. Achúcarro *et al.*, "Inflation: Theory and Observations," 3 2022.

[48] X. Chen and Y. Wang, "Large non-Gaussianities with Intermediate Shapes from Quasi-Single Field Inflation," *Phys. Rev. D* **81**, p. 063511, 2010.

1

[49] D. Baumann and D. Green, "Signatures of Supersymmetry from the Early Universe," *Phys. Rev. D* **85**, p. 103520, 2012.

[50] V. Assassi, D. Baumann, and D. Green, "On Soft Limits of Inflationary Correlation Functions," *JCAP* **11**, p. 047, 2012.

[51] T. Noumi, M. Yamaguchi, and D. Yokoyama, "Effective field theory approach to quasi-single field inflation and effects of heavy fields," *JHEP* **06**, p. 051, 2013.

[52] N. Arkani-Hamed and J. Maldacena, "Cosmological Collider Physics," 3 2015.

[53] H. Lee, D. Baumann, and G. L. Pimentel, "Non-Gaussianity as a Particle Detector," *JHEP* **12**, p. 040, 2016.

[54] X. Chen, M. H. Namjoo, and Y. Wang, "Probing the Primordial Universe using Massive Fields," *Int. J. Mod. Phys. D* **26**(01), p. 1740004, 2016.

[55] E. Komatsu and D. N. Spergel, "Acoustic signatures in the primary microwave background bispectrum," *Phys. Rev. D* **63**, p. 063002, 2001.

[56] Y. Akrami *et al.*, "Planck 2018 results. IX. Constraints on primordial non-Gaussianity," *Astron. Astrophys.* **641**, p. A9, 2020.

[57] W. Sohn, D.-G. Wang, J. R. Fergusson, and E. P. S. Shellard, "Searching for Cosmological Collider in the Planck CMB Data," 4 2024.

[58] J. Silk, "Cosmic black body radiation and galaxy formation," *Astrophys. J.* **151**, pp. 459–471, 1968.

[59] E. Pajer and M. Zaldarriaga, "A New Window on Primordial non-Gaussianity," *Phys. Rev. Lett.* **109**, p. 021302, 2012.

[60] v. Ivezić *et al.*, "LSST: from Science Drivers to Reference Design and Anticipated Data Products," *Astrophys. J.* **873**(2), p. 111, 2019.

[61] R. Laureijs *et al.*, "Euclid Definition Study Report," 10 2011.

[62] D. Spergel *et al.*, "Wide-Field InfrarRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report," 3 2015.

[63] A. Aghamousa *et al.*, "The DESI Experiment Part I: Science,Targeting, and Survey Design," 10 2016.

[64] O. Doré *et al.*, "Cosmology with the SPHEREX All-Sky Spectral Survey," 12 2014.

[65] D. Baumann, A. Nicolis, L. Senatore, and M. Zaldarriaga, "Cosmological Non-Linearities as an Effective Fluid," *JCAP* **07**, p. 051, 2012.

[66] V. Springel, "The Cosmological simulation code GADGET-2," *Mon. Not. Roy. Astron. Soc.* **364**, pp. 1105–1134, 2005.

[67] O. H. E. Philcox, M. M. Ivanov, M. Simonović, and M. Zaldarriaga, "Combining Full-Shape and BAO Analyses of Galaxy Power Spectra: A 1.6\% CMB-independent constraint on H_0," *JCAP* **05**, p. 032, 2020.

[68] O. H. E. Philcox and M. M. Ivanov, "BOSS DR12 full-shape cosmology: ΛCDM constraints from the large-scale galaxy power spectrum and bispectrum monopole," *Phys. Rev. D* **105**(4), p. 043517, 2022.

[69] P. Zhang, G. D'Amico, L. Senatore, C. Zhao, and Y. Cai, "BOSS Correlation Function analysis from the Effective Field Theory of Large-Scale Structure," *JCAP* **02**(02), p. 036, 2022.

[70] G. D'Amico, J. Gleyzes, N. Kokron, K. Markovic, L. Senatore, P. Zhang, F. Beutler, and H. Gil-Marín, "The Cosmological Analysis of the SDSS/BOSS data from the Effective Field Theory of Large-Scale Structure," *JCAP* **05**, p. 005, 2020.

[71] G. D'Amico, Y. Donath, M. Lewandowski, L. Senatore, and P. Zhang, "The BOSS bispectrum analysis at one loop from the Effective Field Theory of Large-Scale Structure," 6 2022.

[72] G. Cabass, M. M. Ivanov, O. H. E. Philcox, M. Simonović, and M. Zaldarriaga, "Constraints on multifield inflation from the BOSS galaxy survey," *Phys. Rev. D* **106**(4), p. 043506, 2022.

[73] G. Cabass, M. M. Ivanov, O. H. E. Philcox, M. Simonović, and M. Zaldarriaga, "Constraints on Single-Field Inflation from the BOSS Galaxy Survey," *Phys. Rev. Lett.* **129**(2), p. 021301, 2022.

[74] G. D'Amico, M. Lewandowski, L. Senatore, and P. Zhang, "Limits on primordial non-Gaussianities from BOSS galaxy-clustering data," 1 2022.

[75] E.-M. Mueller *et al.*, "Primordial non-Gaussianity from the completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey II: measurements in Fourier space with optimal weights," *Mon. Not. Roy. Astron. Soc.* **514**(3), pp. 3396–3409, 2022.

[76] G. Cabass, M. M. Ivanov, O. H. E. Philcox, M. Simonovic, and M. Zaldarriaga, "Constraining single-field inflation with MegaMapper," *Phys. Lett. B* **841**, p. 137912, 2023.

[77] D. Braganca, Y. Donath, L. Senatore, and H. Zheng, "Peeking into the next decade in Large-Scale Structure Cosmology with its Effective Field Theory," 7 2023.

[78] D. J. Schlegel *et al.*, "The MegaMapper: A Stage-5 Spectroscopic Instrument Concept for the Study of Inflation and Dark Energy," 9 2022.

[79] S. Furlanetto, S. P. Oh, and F. Briggs, "Cosmology at Low Frequencies: The 21 cm Transition and the High-Redshift Universe," *Phys. Rept.* **433**, pp. 181–301, 2006.

[80] G. Mellema *et al.*, "Reionization and the Cosmic Dawn with the Square Kilometre Array," *Exper. Astron.* **36**, pp. 235–318, 2013.

1

[81] Z. S. Ali *et al.*, "PAPER-64 Constraints on Reionization: The 21cm Power Spectrum at $z =$8.4," *Astrophys. J.* **809**(1), p. 61, 2015.

[82] D. R. DeBoer *et al.*, "Hydrogen Epoch of Reionization Array (HERA)," *Publ. Astron. Soc. Pac.* **129**(974), p. 045001, 2017.

[83] A. Weltman *et al.*, "Fundamental physics with the Square Kilometre Array," *Publ. Astron. Soc. Austral.* **37**, p. e002, 2020.

[84] Z. Abdurashidova *et al.*, "HERA Phase I Limits on the Cosmic 21 cm Signal: Constraints on Astrophysics and Cosmology during the Epoch of Reionization," *Astrophys. J.* **924**(2), p. 51, 2022.

[85] E. de Lera Acedo *et al.*, "The REACH radiometer for detecting the 21-cm hydrogen signal from redshift z $\approx$ 7.5–28," *Nature Astron.* **6**(7), p. 998, 2022.

[86] J. D. Bowman, A. E. E. Rogers, R. A. Monsalve, T. J. Mozdzen, and N. Mahesh, "An absorption profile centred at 78 megahertz in the sky-averaged spectrum," *Nature* **555**(7694), pp. 67–70, 2018.

[87] P. Bull, P. G. Ferreira, P. Patel, and M. G. Santos, "Late-time cosmology with 21cm intensity mapping experiments," *Astrophys. J.* **803**(1), p. 21, 2015.

[88] R. Ansari *et al.*, "Inflation and Early Dark Energy with a Stage II Hydrogen Intensity Mapping experiment," 10 2018.

[89] U.-L. Pen, L. Staveley-Smith, J. Peterson, and T.-C. Chang, "First Detection of Cosmic Structure in the 21-cm Intensity Field," *Mon. Not. Roy. Astron. Soc.* **394**, p. 6, 2009.

[90] T.-C. Chang, U.-L. Pen, K. Bandura, and J. B. Peterson, "Hydrogen 21-cm Intensity Mapping at redshift 0.8," *Nature* **466**, pp. 463–465, 2010.

[91] K. W. Masui *et al.*, "Measurement of 21 cm brightness fluctuations at z ~ 0.8 in cross-correlation," *Astrophys. J. Lett.* **763**, p. L20, 2013.

[92] M. Amiri *et al.*, "Detection of Cosmological 21 cm Emission with the Canadian Hydrogen Intensity Mapping Experiment," *Astrophys. J.* **947**(1), p. 16, 2023.

[93] M. Amiri *et al.*, "A Detection of Cosmological 21 cm Emission from CHIME in Cross-correlation with eBOSS Measurements of the Lyα Forest," *Astrophys. J.* **963**(1), p. 23, 2024.

[94] S. Paul, M. G. Santos, Z. Chen, and L. Wolz, "A first detection of neutral hydrogen intensity mapping on Mpc scales at $z \approx 0.32$ and $z \approx 0.44$," 1 2023.

[95] J. R. Pritchard and A. Loeb, "21-cm cosmology," *Rept. Prog. Phys.* **75**, p. 086901, 2012.

[96] A. Cooray, "21-cm Background Anisotropies Can Discern Primordial Non-Gaussianity," *Phys. Rev. Lett.* **97**, p. 261301, 2006.

[97] J. Silk, "The limits of cosmology: role of the Moon," *Phil. Trans. A. Math. Phys. Eng. Sci.* **379**, p. 20190561, 2021.

[98] Y. Ali-Haïmoud, P. D. Meerburg, and S. Yuan, "New light on 21 cm intensity fluctuations from the dark ages," *Phys. Rev. D* **89**(8), p. 083506, 2014.

[99] X. Chen, P. D. Meerburg, and M. Münchmeyer, "The Future of Primordial Features with 21 cm Tomography," *JCAP* **09**, p. 023, 2016.

[100] R. Mondal and R. Barkana, "Prospects for precision cosmology with the 21 cm signal from the dark ages," *Nature Astron.* **7**(9), pp. 1025–1030, 2023.

[101] A. Pillepich, C. Porciani, and S. Matarrese, "The bispectrum of redshifted 21-cm fluctuations from the dark ages," *Astrophys. J.* **662**, pp. 1–14, 2007.

[102] J. B. Muñoz, Y. Ali-Haïmoud, and M. Kamionkowski, "Primordial non-gaussianity from the bispectrum of 21-cm fluctuations in the dark ages," *Phys. Rev. D* **92**(8), p. 083508, 2015.

[103] P. D. Meerburg, M. Münchmeyer, J. B. Muñoz, and X. Chen, "Prospects for Cosmological Collider Physics," *JCAP* **03**, p. 050, 2017.

[104] T. Flöss, T. de Wild, P. D. Meerburg, and L. V. E. Koopmans, "The Dark Ages' 21-cm trispectrum," *JCAP* **06**(06), p. 020, 2022.

[105] R. Scoccimarro, E. Sefusatti, and M. Zaldarriaga, "Probing primordial non-Gaussianity with large - scale structure," *Phys. Rev. D* **69**, p. 103513, 2004.

[106] E. Sefusatti, M. Crocce, S. Pueblas, and R. Scoccimarro, "Cosmology and the Bispectrum," *Phys. Rev. D* **74**, p. 023522, 2006.

[107] K. C. Chan and L. Blot, "Assessment of the Information Content of the Power Spectrum and Bispectrum," *Phys. Rev. D* **96**(2), p. 023528, 2017.

[108] M. Biagetti, L. Castiblanco, J. Noreña, and E. Sefusatti, "The covariance of squeezed bispectrum configurations," *JCAP* **09**, p. 009, 2022.

[109] T. Flöss, M. Biagetti, and P. D. Meerburg, "Primordial non-Gaussianity and non-Gaussian covariance," *Phys. Rev. D* **107**(2), p. 023528, 2023.

[110] A. Slosar *et al.*, "Packed Ultra-wideband Mapping Array (PUMA): A Radio Telescope for Cosmology and Transients," *Bull. Am. Astron. Soc.* **51**, p. 53, 2019.

[111] D. Karagiannis, A. Slosar, and M. Liguori, "Forecasts on Primordial non-Gaussianity from 21 cm Intensity Mapping experiments," *JCAP* **11**, p. 052, 2020.

[112] A. Moradinezhad Dizgah, H. Lee, M. Schmittfull, and C. Dvorkin, "Capturing non-Gaussianity of the large-scale structure with weighted skew-spectra," *JCAP* **04**, p. 011, 2020.

[113] E. Massara, F. Villaescusa-Navarro, S. Ho, N. Dalal, and D. N. Spergel, "Using the Marked Power Spectrum to Detect the Signature of Neutrinos in Large-Scale Structure," *Phys. Rev. Lett.* **126**(1), p. 011301, 2021.

[114] G. Valogiannis and C. Dvorkin, "Towards an optimal estimation of cosmological parameters with the wavelet scattering transform," *Phys. Rev. D* **105**(10), p. 103534, 2022.

[115] T. Flöss and P. D. Meerburg, "Improving constraints on primordial non-Gaussianity using neural network based reconstruction," *JCAP* **02**, p. 031, 2024.

[116] F. Villaescusa-Navarro *et al.*, "The Quijote Simulations," **250**, p. 2, Sept. 2020.

[117] S. Cole and G. Efstathiou, "Gravitational lensing of fluctuations in the microwave background radiation," **239**, pp. 195–200, July 1989.

[118] E. V. Linder, "Analysis of gravitationally lensed microwave background anisotropies," **243**, pp. 353–361, Apr. 1990.

[119] U. Seljak, "Gravitational lensing effect on cosmic microwave background anisotropies: A Power spectrum approach," *Astrophys. J.* **463**, p. 1, 1996.

[120] R. B. Metcalf and J. Silk, "Gravitational magnification of the cosmic microwave background," *Astrophys. J.* **489**, pp. 1–6, 1997.

[121] A. Lewis and A. Challinor, "Weak gravitational lensing of the CMB," *Phys. Rept.* **429**, pp. 1–65, 2006.

[122] U. Seljak and C. M. Hirata, "Gravitational lensing as a contaminant of the gravity wave signal in CMB," *Phys. Rev. D* **69**, p. 043005, 2004.

[123] D. Babich and M. Zaldarriaga, "Primordial bispectrum information from CMB polarization," *Phys. Rev. D* **70**, p. 083005, 2004.

[124] W. Coulton, A. Miranthis, and A. Challinor, "Biases to primordial non-Gaussianity measurements from CMB secondary anisotropies," *Mon. Not. Roy. Astron. Soc.* **523**, pp. 825–848, 2023.

[125] W. R. Coulton, P. D. Meerburg, D. G. Baker, S. Hotinli, A. J. Duivenvoorden, and A. van Engelen, "Minimizing gravitational lensing contributions to the primordial bispectrum covariance," *Phys. Rev. D* **101**(12), p. 123504, 2020.

[126] M. Millea, E. Anderes, and B. D. Wandelt, "Sampling-based inference of the primordial CMB and gravitational lensing," *Phys. Rev. D* **102**(12), p. 123542, 2020.

[127] R. Takahashi, T. Hamana, M. Shirasaki, T. Namikawa, T. Nishimichi, K. Osato, and K. Shiroyama, "Full-sky Gravitational Lensing Simulation for Large-area Galaxy Surveys and Cosmic Microwave Background Experiments," *Astrophys. J.* **850**(1), p. 24, 2017.

[128] A. Kalaja, G. Orlando, A. Bowkis, A. Challinor, P. D. Meerburg, and T. Namikawa, "The reconstructed CMB lensing bispectrum," *JCAP* **04**, p. 041, 2023.

[129] G. F. Chew and S. C. Frautschi, "Principle of Equivalence for All Strongly Interacting Particles Within the S Matrix Framework," *Phys. Rev. Lett.* **7**, pp. 394–397, 1961.

[130] G. F. Chew, "S-Matrix Theory of Strong Interactions without Elementary Particles," *Rev. Mod. Phys.* **34**(3), pp. 394–401, 1962.

[131] H. Elvang, "Bootstrap and amplitudes: a hike in the landscape of quantum field theory," *Rept. Prog. Phys.* **84**(7), p. 074201, 2021.

[132] S. Weinberg, "Infrared photons and gravitons," *Phys. Rev.* **140**, pp. B516–B524, 1965.

[133] P. Creminelli, J. Noreña, and M. Simonović, "Conformal consistency relations for single-field inflation," *JCAP* **07**, p. 052, 2012.

[134] K. Hinterbichler, L. Hui, and J. Khoury, "Conformal Symmetries of Adiabatic Modes in Cosmology," *JCAP* **08**, p. 017, 2012.

[135] M. Mirbabayi and M. Simonović, "Weinberg Soft Theorems from Weinberg Adiabatic Modes," 2 2016.

[136] N. Arkani-Hamed, D. Baumann, H. Lee, and G. L. Pimentel, "The Cosmological Bootstrap: Inflationary Correlators from Symmetries and Singularities," *JHEP* **04**, p. 105, 2020.

[137] H. Goodhew, S. Jazayeri, and E. Pajer, "The Cosmological Optical Theorem," *JCAP* **04**, p. 021, 2021.

[138] S. Jazayeri, E. Pajer, and D. Stefanyszyn, "From locality and unitarity to cosmological correlators," *JHEP* **10**, p. 065, 2021.

[139] E. Pajer, "Building a Boostless Bootstrap for the Bispectrum," *JCAP* **01**, p. 023, 2021.

[140] G. Cabass, E. Pajer, D. Stefanyszyn, and J. Supeł, "Bootstrapping large graviton non-Gaussianities," *JHEP* **05**, p. 077, 2022.

[141] G. Cabass, S. Jazayeri, E. Pajer, and D. Stefanyszyn, "Parity violation in the scalar trispectrum: no-go theorems and yes-go examples," *JHEP* **02**, p. 021, 2023.

[142] H. Kawai, D. C. Lewellen, and S. H. H. Tye, "A Relation Between Tree Amplitudes of Closed and Open Strings," *Nucl. Phys. B* **269**, pp. 1–23, 1986.

[143] Z. Bern, J. J. M. Carrasco, and H. Johansson, "New Relations for Gauge-Theory Amplitudes," *Phys. Rev. D* **78**, p. 085011, 2008.

[144] Z. Bern, J. J. M. Carrasco, and H. Johansson, "Perturbative Quantum Gravity as a Double Copy of Gauge Theory," *Phys. Rev. Lett.* **105**, p. 061602, 2010.

1

[145] F. Cachazo, S. He, and E. Y. Yuan, "Scattering Equations and Matrices: From Einstein To Yang-Mills, DBI and NLSM," *JHEP* **07**, p. 149, 2015.

[146] D. de Neeling, D. Roest, and S. Veldmeijer, "Flavour-kinematics duality for Goldstone modes," *JHEP* **10**, p. 066, 2022.

[147] A. R. Fazio, "Cosmological correlators, In–In formalism and double copy," *Mod. Phys. Lett. A* **35**(11), p. 2050076, 2020.

[148] D. Baumann, D. Green, A. Joyce, E. Pajer, G. L. Pimentel, C. Sleight, and M. Taronna, "Snowmass White Paper: The Cosmological Bootstrap," in *Snowmass 2021*, 3 2022.

[149] C. de Rham, G. Gabadadze, and A. J. Tolley, "Resummation of Massive Gravity," *Phys. Rev. Lett.* **106**, p. 231101, 2011.

[150] M. Born and L. Infeld, "Foundations of the new field theory," *Proc. Roy. Soc. Lond. A* **144**(852), pp. 425–451, 1934.

1

Chapter 2

The Dark Ages' 21-cm Trispectrum

Abstract

We investigate tomography of 21-cm brightness temperature fluctuations during the Dark Ages as a probe for constraining primordial non-Gaussianity. We expand the 21-cm brightness temperature effects and derive an improved secondary bispectrum and for the first time derive the secondary trispectrum of 21-cm brightness temperature fluctuations. We then forecast the amount of information available from the Dark Ages to constrain primordial non-Gaussianity, including the imprints of massive particle exchange during inflation and we determine how much signal is lost due to secondary non-Gaussianity. We find that although secondary non-Gaussianity swamps the primordial signal, primordial non-Gaussianity can still be extracted with signal-to-noise ratios that surpass current and future CMB experiments by several orders of magnitude, depending on the experimental setup. Furthermore, we conclude that for the bi- and trispectra of massive particle exchange marginalizing over other primordial shapes affects signal-to-noise ratios more severely than secondary shapes. Baryonic pressure effects turn out to have a negligible impact on our forecasts, even at scales close to the Jeans scale. The results of this chapter reinforce the prospects of 21-cm brightness temperature fluctuations from the Dark Ages as the ultimate probe for primordial non-Gaussianity.

2.1   Introduction

Over the last few decades, inflation has become the leading paradigm for describing the early universe. Even the simplest models of inflation accurately predict the nearly Gaussian initial conditions of the universe and match the precise observations of the Cosmic Microwave Background (CMB) by the Planck satellite [1, 2]. Slight deviations from these Gaussian initial conditions of the early universe, colloquially known as primordial non-Gaussianity (pnG), may be used to further constrain the vast space of inflationary theories [3]. Currently, bounds on the size of pnG by CMB and Large Scale Structure (LSS) observations are consistent with a purely Gaussian distribution of fluctuations, and therefore unable to favor a particular theory of inflation. However, upcoming CMB and LSS experiments are forecast to improve on these bounds and might find a statistically significant deviation from Gaussianity.

The size of pnG can be probed through the higher statistical moments of the distribution of initial conditions. For a purely Gaussian distribution, all information is then contained in the two-point correlation function of the primordial density field. Generally, the size of pnG is determined by the interactions of the inflaton, the scalar field driving the supposed inflationary expansion. However, even in the absence of such *direct* interactions, one expects *indirect* interaction through gravity (i.e. through coupling to the metric). This gravitational interaction sets the minimum amount of pnG that is guaranteed to be present in the initial conditions sourced by inflation and is commonly referred to as the *gravitational floor*. In the simplest theories such as single-field slow-roll inflation, this is the only source of pnG [4, 5], whereas more complicated models (i.e. inflaton self-interactions or multiple fields) could generate considerably larger amounts of non-Gaussianity. Another powerful probe of inflation is the famous Maldacena consistency condition [4, 6], which predicts the vanishing of the squeezed bispectrum for all single-field models of inflation. Hence a non-zero measurement of the squeezed bispectrum would rule out all single-field inflationary models and point toward multi-field dynamics.

Besides being a powerful probe of the physics driving the inflationary expansion, pnG can also contain signatures of additional physics at play during the inflationary era. Massive (spinning) extra fields could have been present, leaving their imprint on the higher order statistics of the initial conditions [7–9]. More specifically, heavy fields induce a characteristic oscillatory shape in the squeezed and collapsed limit of the three- and four-point correlation function of fluctuations respectively. The amplitude and frequency of this oscillation are directly related to the mass, in principle allowing one to probe the particle spectrum during inflation. Such extra fields are naturally present in string theory in which one might ultimately aim to embed the

theory of inflation as well. Hence, we are motivated to study the existence of these extra fields and use inflation as a cosmological particle collider [7]. Although such oscillatory behavior is a clean probe of the particle spectrum, its amplitude is unfortunately severely suppressed proportional to the mass, making it challenging for CMB and LSS experiments to ever measure such an effect.

Clearly, in order to advance precision cosmology to measure pnG, new probes and experiments are necessary. In this chapter, we will study the use of 21-cm fluctuations during the cosmic Dark Ages as the ultimate probe of pnG. The Dark Ages refer to the epoch between recombination ($z \approx 1100$) when neutral hydrogen is formed and the CMB photons are released to free stream and the formation of the first luminous objects (i.e. stars) at $z < 30$. Soon after the CMB photons have been released, they redshift out of the visible wavelengths and the universe becomes truly dark. During this time, a neutral hydrogen gas permeates the universe while occasionally scattering with CMB photons, sometimes exciting the hyper-fine state of the electrons in the neutral hydrogen atoms. When the electron relaxes to its ground state, a photon with a wavelength of 21 centimeters is released. Once the hydrogen gas has cooled sufficiently and its temperature is decoupled from the CMB temperature, this 21-cm signal can be observed in absorption or emission to the background of CMB photons. Higher density regions containing more hydrogen have a brighter 21-cm signal, thereby tracing the matter density field which in turn traces the primordial fluctuations seeded by inflation. In this way, the 21-cm signal during $30 \leq z \leq 100$ can be used as a probe of pnG [10, 11], containing an amount of information that is estimated to be several magnitudes more than that of the CMB and LSS, making it the ultimate probe of the early universe and possibly opening up the cosmological collider [12]. However, several comments are in order. First, the small amplitude of the 21-cm signal makes it hard to measure even the global (mean) brightness temperature, let alone tiny fluctuations around it, making it a serious challenge from a technological and experimental point of view. Secondly, Earth's ionosphere is opaque to the red-shifted 21-cm signal emitted at $z > 30$. Hence, probing this era will require an observatory in space or even on the far side of the moon, where also radio frequency interference (RFI) is minimized [13]. Thirdly, although the 21-cm signal is expected to be a rather pristine tracer of the primordial initial conditions compared to the galaxy distributions that are the target of LSS experiments, it is nevertheless affected by non-linearity (i.e. gravity and peculiar velocity). Both the non-linear dependence of the 21-cm brightness temperature on the initial conditions as well as gravitational effects, therefore, leave their imprint on the distribution of fluctuations, inducing secondary non-Gaussianities that obscure the sought-after primordial non-Gaussianity by several orders of magnitude [11, 14]. A detailed understanding of the 21-cm signal as well as these non-linear effects is warranted in order to reliably

and accurately extract the primordial contribution and learn about inflation.

Secondary non-Gaussianity of the 21-cm brightness temperature during the Dark Ages was addressed previously in [14] for the case of the bispectrum, where it was found to contribute significantly to the observed bispectrum of temperature fluctuations, introducing it as a nuisance that should be marginalized over. In this chapter we will improve on some of the assumptions and simplifications made in [14], in order to more accurately model the 21-cm signal during the Dark Ages. Furthermore, it was recently found that observational sensitivity to the primordial trispectrum can receive an enhanced scaling with respect to the smallest observable scale k_{max} (or ℓ_{max} for CMB surveys) [15]. We will see that this also affects primordial trispectra sourced by massive extra fields, possibly making it of prime observational interest to the cosmological collider. Extracting the primordial trispectrum will require an accurate modeling of the secondary trispectrum, which we derive here for the first time. In an effort to motivate experiments targeted at measuring the 21-cm signal from the Dark Ages, we will determine the total information content of the Dark Ages that can be used to constrain primordial non-Gaussianity. Finally, we forecast the sensitivity of a simple experimental setup, that might be realized in the future.

This chapter is organized as follows. In section 2.2 we review the physics of the 21-cm signal during the Dark Ages. Section 2.3 will cover the non-Gaussian contributions to the statistics of 21-cm fluctuations. Then, in section 2.4, we present Fisher forecasts on the amount of information available to constrain non-Gaussianity from the Dark Ages, as well as the sensitivity of more realistic experimental scenarios. We summarize our conclusions and outlook for future research in section 2.5.

Conventions

For numerical computations, we consider flat ΛCDM cosmology with parameters from *Planck* [16], see Table 2.1.

<div align="center">

ΛCDM parameters

$H_0 = 67.66$	$\Omega_b h^2 = 0.02242$	$\sum m_\nu = 0.06$
$\Omega_k = 0$	$\Omega_c h^2 = 0.11993$	$\tau = 0.0561$
$n_s = 0.9665$	$A_s = 2.1056 \times 10^{-9}$	$r = 0$

</div>

Table 2.1: *Best-fit Planck parameters (specifically, Table 2 of Ref. [16] with TT, TE, EE+lowE+lensing+BAO) used in our numerical computations.*

2.2 21-cm fluctuations during the Dark Ages

In this section, we review the physics that gave rise to the 21-cm brightness fluctuations during the Dark Ages. First, we will restrict the analysis to the background (or global) temperature. Subsequently we will discuss how 21-cm brightness temperature fluctuations arise due to fluctuations in the underlying density fields. In particular, the goal will be to derive how 21-cm fluctuations trace those of the baryonic density and velocity field.

2.2.1 Global 21-cm signal

We define the ratio of the abundance of Hydrogen in the hyperfine singlet ($F = 0$) and triplet state ($F = 1$) via the spin temperature T_s as follows [17]:

$$\frac{n_1}{n_0} \equiv \frac{g_1}{g_0} \, e^{-T_*/T_s}, \tag{2.1}$$

where $T_* \equiv E_{10}/k_{\mathrm{B}}$ is the energy gap between the two hyperfine states and the degeneracies of the levels are $g_{0,1} = 1, 3$. During the Dark Ages, the spin temperature is determined by two processes: collision and radiative transitions between the two states. These processes are described by the rates C_{ij} and R_{ij}, respectively. Specifically, R_{10} encodes spontaneous and stimulated emission whereas R_{01} encodes absorption. The rates C_{01} and C_{10} describe upward and downward transitions between the hyperfine states due to collisions between Hydrogen atoms. We are interested in the Dark Ages, before the first luminous objects are formed, and therefore do not take transitions induced by Lyman-α photons into account (known as the Wouthuysen-Field effect) [17].

During the Dark Ages, the steady-state approximation is very accurate, since $C_{ij} + R_{ij} \gg H$ at all times [18], and the abundances are related via:

$$n_0(C_{01} + R_{01}) = n_1(C_{10} + R_{10}). \tag{2.2}$$

Then, in the limit $T_* \ll T_{\mathrm{gas}}, T_\gamma$, also valid at all times of interest [14], we can write the spin temperature as:

$$T_s = \frac{T_\gamma + Y_{\mathrm{c}} T_{\mathrm{gas}}}{1 + Y_{\mathrm{c}}}, \tag{2.3}$$

where $Y_{\mathrm{c}} \equiv T_* C_{10}/T_{\mathrm{gas}} A_{10}$ and A_{10} the spontaneous decay rate. From the definition of Y_{c}, we find that atomic collisions drive $T_s \to T_{\mathrm{gas}}$, while radiative interactions drive $T_s \to T_\gamma$.

In accordance with the standard convention in the literature, we define the brightness temperature of 21-cm radiation as [14, 18, 19]:

$$T_{21} = \frac{T_s - T_\gamma}{1 + z}\, \tau_{21}, \tag{2.4}$$

where the optical depth for the 21-cm transition, τ_{21}, is given by:

$$\tau_{21} = \frac{3k_{\mathrm{B}}}{32\pi} \frac{T_*}{T_s} n_{\mathrm{HI}} \lambda_{21}^3 \frac{A_{10}}{H + \partial_\parallel v_\parallel}, \tag{2.5}$$

with wavelength $\lambda_{21} = hc/E_{10}$. Note that τ_{21} depends implicitly on the electron fraction via $n_{\mathrm{HI}} \equiv n_{\mathrm{H}}(1 - x_e)$. During the Dark Ages, however, $x_e \sim 10^{-4}$, so that we may approximate $n_{\mathrm{H}} = n_{\mathrm{HI}}$ and the optical depth τ_{21} becomes independent of the free electron fraction. We denote by $\partial_\parallel v_\parallel$ the gradient of the component of the peculiar velocity along the line-of-sight ($v_\parallel$). We define the dimensionless velocity gradient:[1]

$$\delta_v \equiv -\frac{\partial_\parallel v_\parallel}{H}. \tag{2.6}$$

Then, we may write the 21-cm brightness temperature in the following way:

$$T_{21}(T_s, n_{\mathrm{H}}, \delta_v) = \frac{3k_{\mathrm{B}}}{32\pi} \frac{T_*}{T_s} n_{\mathrm{H}} \lambda_{21}^3 \frac{A_{10}}{H(1 - \delta_v)} \frac{T_s - T_\gamma}{1 + z}. \tag{2.7}$$

Notice that the brightness temperature depends explicitly on n_{H} and δ_v and implicitly on T_{gas} via T_s, which in turn depends on n_{H}, making the temperature a non-linear tracer of the underlying density field. For later convenience, we define $\tilde{T}_{21}$ as the brightness temperature in which the dependence on the velocity gradient term δ_v is factored out:

$$\tilde{T}_{21}(T_s, n_{\mathrm{H}}) \equiv T_{21}(T_s, n_{\mathrm{H}}, \delta_v) \times (1 - \delta_v). \tag{2.8}$$

2.2.2 Fluctuations in the 21-cm signal

The final expression for the 21-cm brightness temperature in equation (2.7) shows that it depends on the hydrogen density n_{H}, velocity gradient δ_v, and gas temperature T_{gas} and the photon temperature T_γ. Therefore, fluctuations in the brightness (δT_{21}) are in principle sourced by those in the hydrogen density, velocity gradient, gas, and photon temperature. However, following [14, 19], we will restrict the analysis to sub-horizon scales. On these scales, we assume the photon temperature to be uniform, since their high sound speed ($c_s = c/\sqrt{3}$) results in photon fluctuations being suppressed relative to those in the other fields. In this section, we will find out how 21-cm brightness fluctuations trace those in the hydrogen density, gas temperature, and velocity gradient.

---

[1]In literature, this is sometimes defined without a minus sign (e.g. [19]).

21-cm brightness temperature

We start by considering fluctuations in the density and gas temperature and only include the velocity gradient fluctuations δ_v once we transform to momentum space.[2] That is, we effectively expand $\tilde{T}_{21}$ in terms of fluctuations in n_{H} and T_{gas}. We define fluctuations in the latter by:

$$n_{\mathrm{H}}(t, \mathbf{x}) = \bar{n}_{\mathrm{H}}(t) \Big[1 + \delta_b(t, \mathbf{x}) \Big], \qquad T_{\mathrm{gas}}(t, \mathbf{x}) \equiv \bar{T}_{\mathrm{gas}}(t) \Big[1 + \delta_{\mathrm{T}}(t, \mathbf{x}) \Big], \qquad (2.9)$$

where we have used that $\delta_{\mathrm{H}} = \delta_b$ up to negligible corrections of order $\mathcal{O}(m_e/m_p)$ [19].

Taylor expanding δT_{21} around $\delta_{\mathrm{T}} = \delta_b = 0$ up to cubic order in fluctuations, we find:

$$\begin{aligned}
\delta\tilde{T}_{21}(t, \mathbf{x}) = \; & \mathcal{T}_{\mathrm{T}}\delta_{\mathrm{T}} + \mathcal{T}_b\delta_b \\
& + \mathcal{T}_{\mathrm{TT}}\delta_{\mathrm{T}}^2 + \mathcal{T}_{b\mathrm{T}}\delta_b\delta_{\mathrm{T}} + \mathcal{T}_{bb}\delta_b^2 \\
& + \mathcal{T}_{\mathrm{TTT}}\delta_{\mathrm{T}}^3 + \mathcal{T}_{b\mathrm{TT}}\delta_b\delta_{\mathrm{T}}^2 + \mathcal{T}_{bb\mathrm{T}}\delta_b^2\delta_{\mathrm{T}} + \mathcal{T}_{bbb}\delta_b^3,
\end{aligned} \qquad (2.10)$$

where the coefficients $\mathcal{T}_{b^i\mathrm{T}^j}(t)$ depend only on time. We have numerically solved for the coefficients $\mathcal{T}$ following [19] and show them in Figure 2.1.
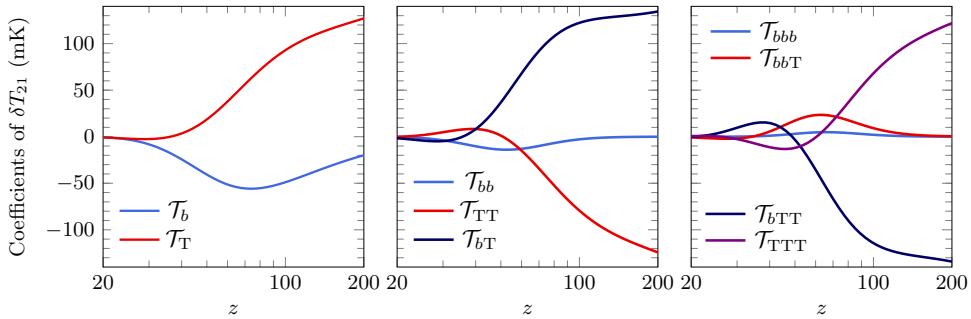
Figure 2.1: *Coefficients $\mathcal{T}$ coupling brightness temperature fluctuations to the gas temperature and baryon density fluctuations. The left, middle, and right panels show the coefficients and linear, quadratic, and cubic order, respectively.*

We can qualitatively understand the behavior of the coefficients $\mathcal{T}$ in the low and high-redshift limits as follows.

---

[2]Since spatial operators such as the gradient transform to simple products in momentum space, it is convenient to consider δ_v only once we have moved to momentum space.

- *Efficient Collisional Coupling.*—At $z \gtrsim 100$, collisions between Hydrogen atoms efficiently couple $T_s = T_{\text{gas}}$. The brightness temperature, in the absence of the velocity gradient term, then depends on n_{H} and T_{gas} as:

$$T_{21} \propto n_{\text{H}} \left(1 - \frac{T_{\gamma}}{T_{\text{gas}}} \right). \tag{2.11}$$

Since $T_{21} \propto n_{\text{H}}$, we find that $\mathcal{T}_b \to \bar{T}_{21}$ and:

$$\mathcal{T}_{bb} \simeq \mathcal{T}_{bbb} \simeq \mathcal{T}_{bbT} \to 0. \tag{2.12}$$

Note also that T_{21} is proportional to the difference $T_{\text{gas}} - T_{\gamma}$ in this regime, which asymptotes to zero in the high redshift limit ($z \gtrsim 200$) where Compton heating efficiently coupling the gas temperature to the photons, such that $\mathcal{T}_b \to 0$. For the coefficients $\mathcal{T}_{\text{T}}$, $\mathcal{T}_{\text{TT}}$ and $\mathcal{T}_{b\text{T}}$ we have:

$$\mathcal{T}_{\text{T}}, \mathcal{T}_{b\text{T}}, \mathcal{T}_{\text{TTT}} \propto \bar{n}_{\text{H}} \frac{\bar{T}_{\gamma}}{\bar{T}_{\text{gas}}}, \qquad \mathcal{T}_{\text{TT}}, \mathcal{T}_{b\text{TT}} \propto -\bar{n}_{\text{H}} \frac{\bar{T}_{\gamma}}{\bar{T}_{\text{gas}}}, \tag{2.13}$$

which yields for $z \gtrsim 100$ the following relations between the coefficients:

$$\mathcal{T}_{\text{T}} \simeq \mathcal{T}_{b\text{T}} \simeq \mathcal{T}_{\text{TTT}} \simeq -\mathcal{T}_{\text{TT}} \simeq -\mathcal{T}_{b\text{TT}}. \tag{2.14}$$

Note that these coefficients do not tend to zero, since they are not suppressed by the factor $T_{\text{gas}} - T_{\gamma}$. Instead, they grow as:

$$T_{21} \propto \frac{n_{\text{H}}}{H} \propto (1+z)^{3/2}, \tag{2.15}$$

during matter domination via the dependence of the optical depth on the ratio n_{H}/H. The above behavior is indeed verified by the numerical solutions in Figure 2.1.

- *Ineffective Collisional Coupling.*—For $z \lesssim 50$, collisions become very inefficient and T_s approaches T_{γ} again, which a small difference:

$$T_s - T_{\gamma} \propto n_{\text{H}} \kappa_{10}^{\text{HH}}(T_{\text{gas}}). \tag{2.16}$$

The brightness temperature then scales approximately as $T_{21} \propto n_{\text{H}}^2$, implying:

$$\mathcal{T}_b \simeq 2\mathcal{T}_{bb} \simeq 2\bar{T}_{21}, \qquad \mathcal{T}_{bbb} \simeq 0. \tag{2.17}$$

As $z \to 0$, the optical depth decreases rapidly and all coefficients tend to zero quickly.

Gas temperature

The evolution equation for the gas temperature T_{gas} can be obtained from the first law of thermodynamics. Neglecting anything other than Compton heating, it reads [19]:

$$\dot{T}_{\text{gas}} - \frac{2}{3}\frac{\dot{n}_{\text{H}}}{n_{\text{H}}}T_{\text{gas}} = \Gamma_{\text{C}}\frac{x_e}{\bar{x}_e}(T_\gamma - T_{\text{gas}}), \tag{2.18}$$

where we have defined the Compton heating rate Γ_{C} as:

$$\Gamma_{\text{C}} \equiv \frac{8\sigma_{\text{T}}a_r T_\gamma^4}{3(1 + x_{\text{He}} + x_e)m_e}\bar{x}_e. \tag{2.19}$$

In the above equation, σ_{T} is the Thomson cross section, a_r is the radiation constant, m_e is the electron mass, $n_{\text{H}} \equiv n_{\text{HI}} + n_p$ is the total density of Hydrogen (both in neutral and ionized form), $x_{\text{He}} \equiv n_{\text{He}}/n_{\text{H}}$ is the Helium fraction, $x_e \equiv n_e/n_{\text{H}}$ is the free electron fraction (and $\bar{x}_e$ its background value). Perturbing the above equation allows us to find a direct relationship between gas temperature fluctuations δ_{T} and baryon density fluctuations δ_b.

To obtain the evolution equation for δ_{T}, one should in principle consistently include the coupling to fluctuations in T_γ, n_{H}, x_e, and x_{He}. To simplify the analysis, we will make the same assumptions as in [19]. As mentioned above, we consider small scales deep inside the horizon ($k_{\text{min}} > 0.01$ Mpc^{-1}), so that photon temperature fluctuations are negligible and we set $T_\gamma = \bar{T}_\gamma$. Secondly, we assume the Helium fraction to be uniform, i.e. we take $x_{\text{He}} = \bar{x}_{\text{He}}$ and neglect any fluctuations. Finally, in the Compton heating rate Γ_{C}, the free electron fraction only enters via the term $1 + x_e + x_{\text{He}}$. During the Dark Ages, $x_e \ll 1$, so that any electron fraction perturbation of Γ_{C} enters through $\bar{x}_e^2\delta_{x_e} \ll \delta_{x_e} \ll 1$. Hence, we can neglect electron fraction perturbations in Γ_{C} such that it only depends on background values $\bar{x}_e, \bar{x}_{\text{He}}$, and $\bar{T}_\gamma$.

The full non-linear evolution of the gas temperature fluctuation $\delta_{\text{T}}(t, \mathbf{x})$ can be obtained directly from equation (2.18) and reads [19]:

$$\dot{\delta}_{\text{T}} - \frac{2}{3}\dot{\delta}_b\frac{1 + \delta_{\text{T}}}{1 + \delta_b} + \frac{\bar{T}_\gamma}{\bar{T}_{\text{gas}}}\Gamma_{\text{C}}\delta_{\text{T}} = \Gamma_{\text{C}}\left[(\bar{T}_\gamma/\bar{T}_{\text{gas}} - 1)\delta_{x_e} + \delta_{x_e}\delta_{\text{T}}\right]. \tag{2.20}$$

We obtain the background evolution of $\bar{T}_{\text{gas}}$ as well as the recombination history $\bar{x}_e$ from CAMB. Note that, as written in the form above, the right-hand side couples the evolution of δ_{T} to the electron fraction fluctuation δ_{x_e}. In [14], the coupling to δ_{x_e} is neglected (i.e. the right-hand side is taken to be zero), resulting in errors of order 10% at linear order during the Dark Ages. In this chapter, we will include the effect

of electron fraction perturbations up to third order. To do so, we require an additional equation describing the evolution of δ_{x_e}.

However, before deriving this additional equation, let us find out when the effect of δ_{x_e} on gas temperature fluctuations is substantial by examining the r.h.s. of equation (2.20). At high redshifts ($z \gtrsim 500$), Compton scattering efficiently couples the gas and CMB temperature, rendering the term proportional to $(\bar{T}_\gamma / \bar{T}_{\text{gas}} - 1)$ vanishingly small. In addition, the strong coupling combined with the fact that we consider scales on which $T_\gamma = \bar{T}_\gamma$ implies $\delta_{\text{T}} = 0$, and hence the second term ($\propto \delta_{x_e} \delta_{\text{T}}$) is negligibly as well. In the low redshift regime ($z \ll 200$), the gas cools adiabatically as $\Gamma_{\text{C}} \ll H$ and fluctuations in the electron fraction also have no effect. However, at intermediate stages, the effect is expected to be non-negligible.

Free electron fraction

Based on the discussion above, we only require an evolution equation for the free electron fraction that is accurate at late times ($z \lesssim 500$), and need not worry about the detailed recombination history at early times [14, 20]. We adapt the effective 3-level model for recombination [21]:

$$\dot{x}_e = -C_{\text{P}}(T_{\text{gas}}, n_{\text{H}}, x_e) \Big[n_{\text{H}} x_e^2 \alpha_{\text{B}}(T_{\text{gas}}) - \beta_{\text{B}}(T_{\text{gas}})(1 - x_e)\, e^{-E_{21}/k_{\text{B}} T_{\text{gas}}} \Big], \quad (2.21)$$

where $E_{21} = 10.2$ eV is the energy difference of the Ly-α transition, α_{B} is the case-B effective recombination coefficient, β_{B} the corresponding photo-ionization rate. Finally, the Peebles factor C_{P} gives the ratio of the effective downward transition rate from the $n = 2$ states to their effective lifetime:

$$C_{\text{P}}(T_{\text{gas}}, n_{\text{H}}, x_e) = \frac{1 + K\Lambda_{2s \to 1s} n_{\text{H}}(1 - x_e)}{1 + K\Lambda_{2s \to 1s} n_{\text{H}}(1 - x_e) + K\beta_{\text{B}} n_{\text{H}}(1 - x_e)}, \quad (2.22)$$

where $K \equiv \lambda_\alpha^3 / 8\pi H$ and $\lambda_\alpha = 121.5$ nm is the Ly-α rest wavelength.

Now, we simplify the evolution equation at late times. For $z \lesssim 500$, the second term in equation (2.21) describing the effect of photo-ionization, is completely negligible compared to the recombination term proportional to α_{B}. In addition, $C_{\text{P}} \to 1$ for $z \lesssim 900$ [19], so that we obtain:

$$\dot{x}_e = -\alpha_{\text{B}} n_{\text{H}} x_e^2, \quad (2.23)$$

to excellent precision. For the recombination coefficient α_{B}, we use the fit [22]:

$$\alpha_{\text{B}}(T_{\text{gas}}) = F \frac{a_\alpha T_4^b}{1 + c T_4^d}, \quad (2.24)$$

where $T_4 \equiv T_{\text{gas}}/10^4$ K, $a_\alpha = 4.309 \times 10^{-19}$ m^3/s, $b = -0.6166$, $c = 0.6703$ and $d = 0.5300$. The Fudge factor $F = 1.14$ is used to calibrate the effective 3-level result to a multi-level atom calculation [20].

To obtain the evolution equation for δ_{x_e}, we perturb equation (2.23). Notice that the evolution of fluctuations in the electron fraction is coupled to those in the gas temperature and the baryon density via $\alpha_{\text{B}}(T_{\text{gas}})$ and n_{H}, respectively. We expand α_{B} in terms of δ_{T} up to cubic order as follows:

$$\alpha_{\text{B}}(T_{\text{gas}}) = \bar{\alpha}_{\text{B}}\left[1 + \sum_{n=1}^{3} A_n \delta_{\text{T}}^n\right], \qquad A_n \equiv \frac{1}{n!} \frac{\bar{T}_{\text{gas}}^n}{\bar{\alpha}_{\text{B}}} \frac{\partial^n \bar{\alpha}_{\text{B}}}{\partial \bar{T}_{\text{gas}}^n}, \tag{2.25}$$

where $\bar{\alpha}_{\text{B}} \equiv \alpha_{\text{B}}(\bar{T}_{\text{gas}})$. Upon defining $\Gamma_{\text{R}} \equiv \bar{\alpha}_{\text{B}} \bar{n}_{\text{H}} \bar{x}_e$ as the background recombination rate, we obtain the evolution of electron fraction perturbations up to terms of cubic order in the baryon, gas temperature, and electron fraction fluctuations:

$$\begin{aligned}
\dot{\delta}_{x_e} = -\Gamma_{\text{R}}\Big[& \delta_{x_e} + A_1 \delta_{\text{T}} + \delta_b \\
& + \delta_{x_e}^2 + A_2 \delta_{\text{T}}^2 + 2A_1 \delta_{x_e} \delta_{\text{T}} + A_1 \delta_{\text{T}} \delta_b + 2\delta_{x_e} \delta_b \\
& + A_3 \delta_{\text{T}}^3 + 2A_2 \delta_{x_e} \delta_{\text{T}}^2 + A_1 \delta_{x_e}^2 \delta_{\text{T}} + A_2 \delta_{\text{T}}^2 \delta_b + \delta_{x_e}^2 \delta_b + 2A_1 \delta_{x_e} \delta_{\text{T}} \delta_b \Big].
\end{aligned} \tag{2.26}$$

2.2.3 Perturbative analysis of fluctuations

Recall that the goal of this section is to find the relation between δT_{21} and the underlying baryonic density and velocity field. At this point, δT_{21} is still a function of fluctuations in the gas temperature as well. However, δ_{T} can be traded effectively for δ_b by recognizing that gas temperature fluctuations trace fluctuations in the density field. The physical mechanism behind this tracing relationship is simple and may be explained as follows. Consider an overdense region in the gas ($\delta_b > 0$). Due to the higher density, the thermal motion of the particles in the gas is increased and hence the temperature will increase as well: δ_{T} traces δ_b.

In order to find the tracing relationship between δ_{T} and δ_b, we start by writing the baryonic density contrast perturbatively, up to cubic order:

$$\delta_b = \sum_{n=1}^{3} \delta_b^{(n)} \equiv \sum_{n=1}^{3} \delta_n, \tag{2.27}$$

where we assume $\delta_b^{(n)} = \mathcal{O}\left[\left(\delta_b^{(1)}\right)^n\right]$. In the second equality, we defined $\delta_n \equiv \delta_b^{(n)}$ for notational brevity. We assume that the time-dependence of the baryons is iden-

tical to that of CDM, i.e. $\delta_b^{(n)} \propto a^{(n)}(t)$, which implies $\dot{\delta}_b^{(n)} = nH\delta_b^{(n)}$.[3] Similarly, we expand $\delta_{\rm T}$ and δ_{x_e} to cubic order as well:[4]

$$\delta_{x_e}(t, \mathbf{x}) = \sum_{n=1}^{3} \delta_n^x(t, \mathbf{x}), \qquad \delta_{\rm T}(t, \mathbf{x}) = \sum_{n=1}^{3} \delta_n^{\rm T}(t, \mathbf{x}). \qquad (2.28)$$

At each order, the evolution equations for the perturbations in the temperature and electron fraction can be obtained by inserting the above expansions into equation (2.20) and equation (2.26), respectively. In Appendix 2.B we provide the resulting evolution equations and show how they can be used to solve for the coupling coefficients, relating $\delta_{\rm T}$ and δ_{x_e} to δ_b up to third order:

$$\delta_{\rm T}(z, \mathbf{x}) = \mathcal{C}_{1,1}^{\rm T}(z)\, \delta_1 + \mathcal{C}_{2,1}^{\rm T}(z)\, \delta_1^2 + \mathcal{C}_{2,2}^{\rm T}(z)\, \delta_2 + \mathcal{C}_{3,1}^{\rm T}(z)\, \delta_1^3 + \mathcal{C}_{3,2}^{\rm T}(z)\, \delta_1\delta_2 + \mathcal{C}_{3,3}^{\rm T}(z)\, \delta_3,$$
$$(2.29)$$

$$\delta_x(z, \mathbf{x}) = \mathcal{C}_{1,1}^{x}(z)\, \delta_1 + \mathcal{C}_{2,1}^{x}(z)\, \delta_1^2 + \mathcal{C}_{2,2}^{x}(z)\, \delta_2 + \mathcal{C}_{3,1}^{x}(z)\, \delta_1^3 + \mathcal{C}_{3,2}^{x}(z)\, \delta_1\delta_2 + \mathcal{C}_{3,3}^{x}(z)\, \delta_3,$$
$$(2.30)$$

where the coupling coefficients are functions of redshift only. In the coefficients $\mathcal{C}_{n,m}$, n denotes the total order of the combination of density perturbations it multiplies, and $m = 1, \ldots, n$ labels the different coefficients at each order n. Note that $\mathcal{C}_{n,n}$ is the coefficient coupling to density perturbation δ_n, whereas $\mathcal{C}_{n,m\neq n}$ couple to the product of lower order density fields with combined order n. At first order, we will often write $\mathcal{C}_1 \equiv \mathcal{C}_{1,1}$ for notational simplicity.

Evolution of Coupling Coefficients

In Figure 2.2, we show the numerical solutions for the gas coupling coefficients $\mathcal{C}_{n,m}^{\rm T}$. The solid lines show the solutions including the effect of electron fraction perturbations. The dotted lines exclude their effect by setting $\mathcal{C}_{n,m}^{x} \equiv 0$ in all evolution equations. It is insightful to examine the numerical results at early (high-z) and late times (low-z), where the numerical analysis can be compared with analytic results:

- *High redshift limit*—At high redshifts, $z \gtrsim 500$, Compton heating is efficient ($\Gamma_{\rm C}/H \gg 1$) and keeps the matter and photons in equilibrium so that $T_\gamma = T_{\rm gas}$. Since we consider small scales on which we neglect photon temperature fluctuations, we therefore expect no gas temperature fluctuations. This is indeed verified by the solutions for $\mathcal{C}_{n,m}^{\rm T}$, which all tend to zero at high redshifts.

---

[3]Note that we *do not* assume the *spatial* dependence of baryons and CDM to be identical. In fact, we will include the spatial dependence due to baryonic pressure, in contrast to [14].

[4]In the notation below, the superscripts T and x are labels, not powers.
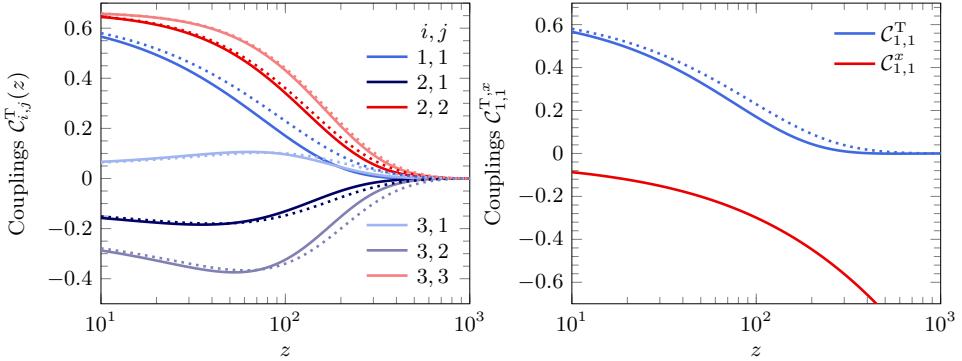
Figure 2.2: *Left Panel: Coupling coefficients $\mathcal{C}_{n,m}^{\mathrm{T}}$ as a function of redshift. The dotted lines show the result in case electron fraction fluctuations are neglected (i.e. $\mathcal{C}_{n,m}^{x} \equiv 0$). Right Panel: First-order coefficients as a function of redshift. Again, the dotted line shows $\mathcal{C}_{1}^{\mathrm{T}}$ in the absence of electron fraction fluctuations.*

Only when the gas temperature starts to decouple from the photon temperature at $z \lesssim 500$ due to adiabatic cooling, do fluctuations in the former start growing.

- *Low redshift limit*—At low redshifts, Compton heating becomes completely inefficient, $\Gamma_{\mathrm{C}}/H \ll 1$, and adiabatic cooling completely determines the gas temperature:

$$\dot{T}_{\mathrm{gas}} = \frac{2}{3}\frac{\dot{n}_{\mathrm{H}}}{n_{\mathrm{H}}}T_{\mathrm{gas}}, \tag{2.31}$$

which is just equation (2.18) in the absence of the Compton heating term (proportional to Γ_{C}). From the above equation, we easily obtain:

$$T_{\mathrm{gas}} \propto n_{\mathrm{H}}^{2/3}, \tag{2.32}$$

characterizing a gas that is cooling adiabatically. The above relationship between the gas temperature and Hydrogen density yields the following relationship between δ_{T} and δ_b:

$$\begin{aligned}
\delta_{\mathrm{T}} &= \frac{2}{3}\delta_b - \frac{1}{9}\delta_b^2 + \frac{4}{81}\delta_b^3 + \mathcal{O}(\delta_b^4) \\
&= \frac{2}{3}(\delta_1 + \delta_2 + \delta_3) - \frac{1}{9}(\delta_1^2 + 2\delta_1\delta_2) + \frac{4}{81}\delta_1^3 + \mathcal{O}(\delta_1^4),
\end{aligned} \tag{2.33}$$

where we have expanded δ_b up the third order in the second line. Comparing to equation (2.92), we find that the coefficients asymptote to constant values in

the adiabatic or low redshift regime:

$$\mathcal{C}_1^{\mathrm{T}} = \mathcal{C}_{2,2}^{\mathrm{T}} = \mathcal{C}_{3,3}^{\mathrm{T}} \to 2/3, \qquad \mathcal{C}_{2,1}^{\mathrm{T}} \to -1/9, \qquad \mathcal{C}_{3,1}^{\mathrm{T}} \to 4/81, \qquad \mathcal{C}_{3,2}^{\mathrm{T}} \to -2/9, \tag{2.34}$$

which is indeed verified by the numerical solution.

Below, we will first discuss the case in which electron fraction perturbations are neglected, and make contact with results obtained in literature. Then, we discuss the effect of electron fluctuations and provide a physical interpretation of their effect at first order.

As mentioned above, excluding the effect of electron fraction fluctuations leads to the dotted curves for the gas temperature coefficients. Our results agree with those of [14], in which the gas temperature coefficients were computed up to second order and in the absence of electron fluctuations. We find excellent agreement for the coefficients $\mathcal{C}_1^{\mathrm{T}}$, $\mathcal{C}_{2,1}^{\mathrm{T}}$ and $\mathcal{C}_{2,2}^{\mathrm{T}}$, which are called C_1, C_2 and C_2' respectively in [14] and shown in their Figure 1. In addition, the first-order coupling coefficient agrees with the result obtained neglecting electron fraction fluctuations in [11], where it is called g_1 (see their Figure 4).

Comparing the solid and dotted curves in Figure 2.2, we find that the effect of electron fraction perturbations is most pronounced at linear order, i.e. for $\mathcal{C}_1^{\mathrm{T}}$. At $z = 30, 50$ and 100, the solution excluding electron fraction perturbations is larger by $8\%, 14\%$ and 34%, respectively. Our results confirm previous claims that neglecting δ_{x_e} leads to errors of order 10% for the linear evolution during the Dark Ages [14]. However, we also note that the effect of δ_{x_e} rapidly becomes more significant at higher redshift during the Dark Ages, and should therefore be accounted for in a detailed analysis. At second and third order, the effect of electron fraction perturbations is smaller but still starts to exceed the 10% level at $z \gtrsim 100$, before decreasing again as $z \to 1100$.

In the low redshift limit, the solutions including and excluding δ_{x_e} agree again, which can be understood as follows. Due to the expansion of the universe, the Hydrogen gas becomes diluted at low redshift, rendering the recombination rate small: $\Gamma_{\mathrm{R}}/H \ll 1$. The evolution equations for $\mathcal{C}_{n,m}^x$ then reduce to the form:

$$\frac{d\mathcal{C}_{n,m}^x}{da} \propto -\frac{\mathcal{C}_{n,m}^x}{a}, \tag{2.35}$$

implying that the electron fraction coefficients decay as $\mathcal{C}_{n,m}^x \propto a^{-1}$ at low redshift. In turn, the effect of electron fraction perturbations on the evolution of $\mathcal{C}_{n,m}^{\mathrm{T}}$ becomes small at low redshifts, and the dotted and solid curves become identical. Below, we

will discuss the effect of δ_{x_e} in more detail at first order.

At linear order, the effect of electron fraction fluctuations on the evolution of gas temperature fluctuations can be understood in an intuitive way from the physical principles at play [20]. Consider a region with a density that is higher than average, so that $\delta_1 > 0$. In such a region, Hydrogen density is higher than average, so the local recombination rate is slightly higher, and hence more recombinations occur. As a consequence, the electron fraction will be lower than average, resulting in $\delta_1^x < 0$ and reflected by $\mathcal{C}_1^x < 0$. In turn, this reduces the coupling of the gas temperature to the photon temperature via Compton scattering, since there are fewer free electrons available to sustain the coupling. At background level, T_{gas} is below T_γ during the Dark Ages due to adiabatic cooling. In an overdense region, the reduced coupling between T_γ and T_{gas} makes the latter even smaller relative to the background difference. This results in δ_1^{T} (or equivalently $\mathcal{C}_1^{\text{T}}$) being smaller than it would have been if electron fraction fluctuations were neglected.

2.2.4 Coupling 21-cm fluctuations to density and velocity fluctuations

At this point, we are in the position to expand 21-cm fluctuations into fluctuations δ_b and δ_v, as we have uniquely related temperature fluctuations to density fluctuations. We will make this expansion explicit in real space first. Subsequently, we transform to momentum space, where we can relate δ_v to the velocity divergence θ_b.

Real Space

Now we have all the ingredients to compute the coupling between the 21-cm brightness temperature and density fluctuations up to third order. We will denote these couplings by $\alpha_{n,m}$, and they are defined in relation to $\delta\tilde{T}_{21}$ as:

$$\delta\tilde{T}_{21}(z,\mathbf{x}) = \alpha_{1,1}(z)\delta_1 + \alpha_{2,1}(z)\delta_1^2 + \alpha_{2,2}(z)\delta_2 + \alpha_{3,1}(z)\delta_1^3 + \alpha_{3,2}(z)\delta_1\delta_2 + \alpha_{3,3}(z)\delta_3. \tag{2.36}$$

Using equation (2.29) and equation (2.30), we can write the couplings $\alpha_{n,m}$ explicitly in terms of the $\mathcal{T}$ coefficients and the couplings $\mathcal{C}_{n,m}^{\text{T}}$, yielding:

$$\alpha_{n,n} \equiv \mathcal{T}_{\text{T}}\mathcal{C}_{n,n}^{\text{T}} + \mathcal{T}_b, \tag{2.37}$$

$$\alpha_{2,1} \equiv \mathcal{T}_{\text{T}}\mathcal{C}_{2,1}^{\text{T}} + \mathcal{T}_{\text{TT}}[\mathcal{C}_1^{\text{T}}]^2 + \mathcal{T}_{b\text{T}}\mathcal{C}_1^{\text{T}} + \mathcal{T}_{bb}, \tag{2.38}$$

$$\alpha_{3,1} \equiv \mathcal{T}_{\text{T}}\mathcal{C}_{3,1}^{\text{T}} + 2\mathcal{T}_{\text{TT}}\mathcal{C}_1^{\text{T}}\mathcal{C}_{2,1}^{\text{T}} + \mathcal{T}_{b\text{T}}\mathcal{C}_{2,1}^{\text{T}} + \mathcal{T}_{\text{TTT}}[\mathcal{C}_1^{\text{T}}]^3 + \mathcal{T}_{b\text{TT}}[\mathcal{C}_1^{\text{T}}]^2 + \mathcal{T}_{bb\text{T}}\mathcal{C}_1^{\text{T}} + \mathcal{T}_{bbb}, \tag{2.39}$$

$$\alpha_{3,2} \equiv \mathcal{T}_{\text{T}}\mathcal{C}_{3,2}^{\text{T}} + 2\mathcal{T}_{\text{TT}}\mathcal{C}_1^{\text{T}}\mathcal{C}_{2,2}^{\text{T}} + \mathcal{T}_{b\text{T}}(\mathcal{C}_{2,2}^{\text{T}} + \mathcal{C}_1^{\text{T}}) + 2\mathcal{T}_{bb}. \tag{2.40}$$

We plot the couplings in Figure 2.3 over the redshift range $10 - 1000$. Recall that in the adiabatic limit $\mathcal{C}_{n,n}^{\mathrm{T}} \to 2/3$ so that the couplings $\alpha_{n,n}$ become identical, which is indeed the case for $z \lesssim 50$. In addition, note that $\mathcal{C}_{3,2}^{\mathrm{T}} \simeq 2\mathcal{C}_{2,1}^{\mathrm{T}}$ and $\mathcal{C}_{2,2}^{\mathrm{T}} \simeq \mathcal{C}_{1}^{\mathrm{T}}$ in the adiabatic limit, such that $\alpha_{3,2} \simeq 2\alpha_{2,1}$, which is indeed the case for $z \lesssim 30$. Note that in the approach above, we have effectively traded the dependence on δ_{T} for dependence on δ_b. Ultimately, we are able to do so by using the first law of thermodynamics.

So far, we have worked with $\delta\tilde{T}_{21}$, i.e. we ignored the fluctuations δ_v. To include the perturbations in δ_v, we use the relation $T_{21} = \tilde{T}_{21}/(1 - \delta_v)$ and expand δ_v to third order, i.e. $\delta_v = \delta_v^{(1)} + \delta_v^{(2)} + \delta_v^{(3)}$.[5] Assuming $|\delta_v| \ll 1$ and using a geometric series expansion, we obtain:

$$\delta T_{21}(z, \mathbf{x}) = \bar{T}_{21}(z)\Big[\delta_v^{(1)} + \big[\delta_v^{(1)}\big]^2 + \delta_v^{(2)} + \big[\delta_v^{(1)}\big]^3 + 2\delta_v^{(1)}\delta_v^{(2)} + \delta_v^{(3)}\Big](z, \mathbf{x})$$
$$+ \alpha_{1,1}(z)\Big[1 + \delta_v^{(1)} + \delta_v^{(2)} + \big[\delta_v^{(1)}\big]^2\Big]\delta_b^{(1)}(z, \mathbf{x})$$
$$+ \alpha_{2,1}(z)\Big[1 + \delta_v^{(1)}\Big]\big[\delta_b^{(1)}\big]^2(z, \mathbf{x}) + \alpha_{2,2}(z)\Big[1 + \delta_v^{(1)}\Big]\delta_b^{(2)}(z, \mathbf{x})$$
$$+ \alpha_{3,1}(z)\big[\delta_b^{(1)}(z, \mathbf{x})\big]^3 + \alpha_{3,2}(z)\big[\delta_b^{(1)}\delta_b^{(2)}\big](z, \mathbf{x}) + \alpha_{3,3}(z)\delta_b^{(3)}(z, \mathbf{x}).$$
$$\text{(2.41)}$$

Momentum Space

In the end, we wish to obtain correlation functions of 21-cm fluctuations in momentum space. Therefore, we transform the above expression for $\delta T_{21}(z, \mathbf{x})$ to momentum space. To do so we have to transform all different components separately. Products of perturbations in real space, e.g. the square of $\delta_b^{(1)}$, become convolutions in momentum space. In addition, the transform of δ_v can be related directly to the transform of the velocity divergence $\theta_b \equiv \nabla \cdot \mathbf{v}_b$. The Fourier transform of δ_v is written in terms of that of θ_b as:

$$\delta_v(z, \mathbf{k}) = -\mu^2(\mathbf{k})\frac{\theta_b(z, \mathbf{k})}{\mathcal{H}}, \qquad \text{(2.42)}$$

where $\mathcal{H} = aH$ and $\mu(\mathbf{k}) \equiv \hat{\mathbf{n}} \cdot \mathbf{k}/k = k_{\parallel}/k$ is the cosine of the angle between the mode $\mathbf{k}$ and the line-of-sight $\mathbf{n}$. At first order, this expression can be simplified even

---

[5]Although the fluctuations δ_v appear similar to *redshift space distortions* (RSD), they come from the velocity dependence of the optical depth and hence are of a qualitative different nature. In principle one should also include RSDs separately when fully modeling the 21-cm signal, introducing the linear growth rate $f = d\ln D_+/d\ln a$ as an additional parameter to marginalize over. We have chosen to ignore RSDs in our forecast as they do not introduce additional shape dependence and therefore overlap with the primordial shapes.
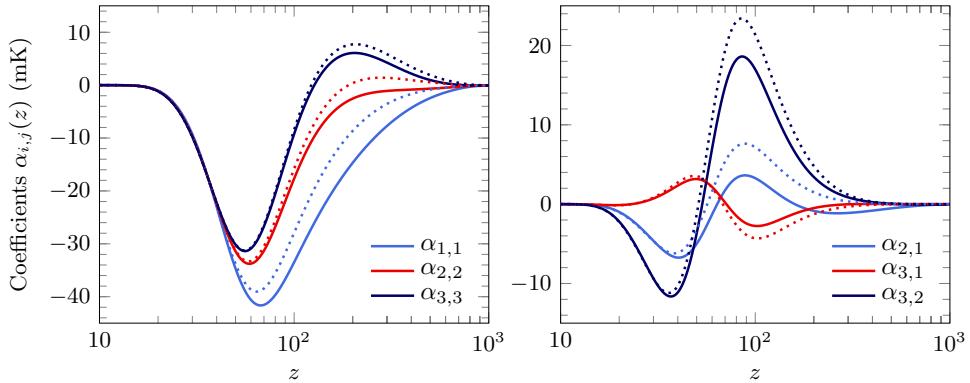
Figure 2.3: *Coupling coefficients $\alpha_{n,n}$ (left panel) and $\alpha_{n,m\neq n}$ (right panel) as function of redshift. Again, the dotted lines exclude the effect of electron fluctuations.*

further by invoking the continuity equation, to obtain

$$\delta_v^{(1)}(z,\mathbf{k}) = \mu^2(\mathbf{k})\delta_b^{(1)}(z,\mathbf{k}). \tag{2.43}$$

Note that in transforming to momentum space, the angle between the line-of-sight and the Fourier mode is encoded in $\mu^2(\mathbf{k})$. This leads to a distinct difference between coupling coefficients in momentum space and those in real space ($\bar{T}_{21}$ and $\alpha_{i,j}$); the former depends on the mode $\mathbf{k}$ and time, whereas the latter depend on only on time and not on $\mathbf{x}$.

Now, we have all the relevant ingredients to compute the Fourier transform of the brightness fluctuation δT_{21}. At each order, we couple the brightness perturbations to those in the baryon density and velocity divergence via the coefficients $c_j^{(i)}(z,\mathbf{k},\mathbf{q})$, of which some explicitly depend on the principal mode $\mathbf{k}$ and internal modes $\mathbf{q}$ resulting from convolution integrals via the angle μ as explained above. The label i denotes the total order of the quantity the coefficient couples to, and j labels all different couplings at order i. Up to third order, the expansion of $\delta T_{21}(z,\mathbf{k})$ is given by:

$$\delta T_{21}(z,\mathbf{k}) = \delta T_{21}^{(1)}(z,\mathbf{k}) + \delta T_{21}^{(2)}(z,\mathbf{k}) + \delta T_{21}^{(3)}(z,\mathbf{k}), \tag{2.44}$$

with $\delta T_{21}^{(i)}(z, \mathbf{k})$ given in terms of the coefficients $c_j^{(i)}$ as:

$$\delta T_{21}^{(1)}(z, \mathbf{k}) = c_1^{(1)}(z, \mathbf{k})\delta_b^{(1)}(\mathbf{k})$$

$$\delta T_{21}^{(2)}(z, \mathbf{k}) = c_1^{(2)}(z)\delta_b^{(2)}(\mathbf{k}) - \frac{1}{\mathcal{H}}c_2^{(2)}(z, \mathbf{k})\theta_b^{(2)}(\mathbf{k})$$

$$+ \int_{\mathbf{q}_1} c_3^{(2)}(z, \mathbf{k}, \mathbf{q}_1)\delta_b^{(1)}(\mathbf{q}_1)\delta_b^{(1)}(\mathbf{k} - \mathbf{q}_1)$$

$$\delta T_{21}^{(3)}(z, \mathbf{k}) = c_1^{(3)}(z)\delta_b^{(3)}(\mathbf{k}) - \frac{1}{\mathcal{H}}c_2^{(3)}(z, \mathbf{k})\theta_b^{(3)}(\mathbf{k})$$

$$+ \int_{\mathbf{q}_1 \mathbf{q}_2} c_3^{(3)}(z, \mathbf{k}, \mathbf{q}_1, \mathbf{q}_2)\delta_b^{(1)}(\mathbf{q}_1)\delta_b^{(1)}(\mathbf{q}_2)\delta_b^{(1)}(\mathbf{k} - \mathbf{q}_{12})$$

$$- \frac{1}{\mathcal{H}} \int_{\mathbf{q}_1} c_4^{(3)}(\mathbf{k}, \mathbf{q}_1)\theta_b^{(2)}(\mathbf{q}_1)\delta_b^{(1)}(\mathbf{k} - \mathbf{q}_1)$$

$$+ \int_{\mathbf{q}_1} c_5^{(3)}(\mathbf{k}, \mathbf{q}_1)\delta_b^{(2)}(\mathbf{q}_1)\delta_b^{(1)}(\mathbf{k} - \mathbf{q}_1), \tag{2.45}$$

where we have suppressed the time dependence in $\delta_b^{(n)}$ and $\theta_b^{(n)}$ for brevity. We explicitly excluded the factors of $1/\mathcal{H}$ from the definition of the couplings $c_j^{(i)}$, so that the latter all have the dimension of temperature. The couplings are a function of $\alpha_{i,j}$, $\bar{T}_{21}$ and $\mu(\mathbf{k})$ and read:

$$c_1^{(1)} \equiv \alpha_{1,1}(z) + \bar{T}_{21}(z)\mu^2(\mathbf{k}),$$

$$c_1^{(2)} \equiv \alpha_{2,2}(z),$$

$$c_2^{(2)} \equiv \bar{T}_{21}\mu^2(\mathbf{k}),$$

$$c_3^{(2)} \equiv \bar{T}_{21}(z)\mu^2(\mathbf{q})\mu^2(\mathbf{k} - \mathbf{q}) + \alpha_{1,1}\mu^2(\mathbf{q}) + \alpha_{2,1}(z),$$

$$c_1^{(3)} \equiv \alpha_{3,3}(z),$$

$$c_2^{(3)} \equiv \bar{T}_{21}(z)\mu^2(\mathbf{k}),$$

$$c_3^{(3)} \equiv \bar{T}_{21}(z)\mu^2(\mathbf{q}_1)\mu^2(\mathbf{q}_2)\mu^2(\mathbf{k} - \mathbf{q}_{12}) + \alpha_{1,1}(z)\mu^2(\mathbf{q}_1)\mu^2(\mathbf{q}_2) + \alpha_{2,1}(z)\mu^2(\mathbf{q}_1)$$

$$+ \alpha_{3,1}(z),$$

$$c_4^{(3)} \equiv 2\bar{T}_{21}(z)\mu^2(\mathbf{q})\mu^2(\mathbf{k} - \mathbf{q}) + \alpha_{1,1}(z)\mu^2(\mathbf{q}),$$

$$c_5^{(3)} \equiv \alpha_{2,2}(z)\mu^2(\mathbf{k} - \mathbf{q}) + \alpha_{3,2}(z). \tag{2.46}$$

Note that since $\mu(\mathbf{k}) = \mu(-\mathbf{k})$, all coefficients are symmetric under inverting all momentum arguments, e.g. $c_3^{(3)}(\mathbf{k}, \mathbf{q}_1, \mathbf{q}_2) = c_3^{(3)}(-\mathbf{k}, -\mathbf{q}_1, -\mathbf{q}_2)$. The first order coupling, $c_1^{(1)}$, actually comprises two couplings, to $\delta_b^{(1)}$ and $\theta_b^{(1)}$, respectively. However, using equation (2.43) allows us to write down one overall coefficient coupling to $\delta_b^{(1)}$. In summary, we have found a direct relationship between δT_{21} and δ_b to third order

in perturbations.

We are now able to express the statistics of 21-cm temperature fluctuations in terms of the underlying tracer field δ_b, which in turn traces primordial fluctuations, e.g. the power spectrum:

$$
\begin{aligned}
P_{\delta T_{21}}(\mathbf{k}_1) = \langle \delta T_{21}(\mathbf{k}_1)\delta T_{21}(-\mathbf{k}_1)\rangle' &= \left(c_1^{(1)}(\mathbf{k}_1)\right)^2 \langle \delta_b(\mathbf{k}_1)\delta_b(-\mathbf{k}_1)\rangle' \\
&= \left(c_1^{(1)}(\mathbf{k}_1)\mathcal{M}_b(\mathbf{k}_1)\right)^2 P_\zeta(\mathbf{k}_1)
\end{aligned}
\tag{2.47}
$$

where $\mathcal{M}_b$ is the linear transfer function of baryon fluctuations (obtainable through e.g. CAMB) and the time (redshift) dependence of the prefactor is implicit.

2.3 Non-Gaussianity of 21-cm brightness temperature anisotropies

In this section, we review the non-Gaussian contributions to the statistics of 21-cm brightness temperature fluctuations during the Dark Ages. First, we will go over the details of primordial non-Gaussianity and present templates that can be used to extract imprints of inflation from 21-cm data. Subsequently, we discuss the generation of secondary non-Gaussianity in the 21-cm signal due to the non-linear evolution of fluctuations, such as gravitational collapse.

2.3.1 Primordial non-Gaussianity

As was mentioned in the introduction, inflation naturally seeds structure formation through the generation of quantum fluctuations. Over time, higher-density regions collapse under gravitational attraction, forming stars and eventually galaxies. The statistics of the initial conditions set by inflation contain a wealth of information about the physics driving the accelerated expansion, as well as any extra particles present during this period. Primordial (scalar) fluctuations are captured by the gauge invariant quantity ζ with mean zero. Generally, the N-point statistical correlation function of primordial fluctuations is given by:[6]

$$
\langle \zeta_{\mathbf{k}_1}\zeta_{\mathbf{k}_2}...\zeta_{\mathbf{k}_N}\rangle = (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + ... + \mathbf{k}_N)F_\zeta^{(N)}(\mathbf{k}_1, \mathbf{k}_2, ..., \mathbf{k}_N)
\tag{2.48}
$$

---

[6]The number of degrees of freedom of the N-point correlation function is reduced by homogeneity and isotropy to be $3(N-2)$.

where $F_\zeta^{(N)}$ encodes the size and *shape* of the N-point correlation as a function of the momentum configuration. The lowest order statistic is known as the power spectrum:[7]

$$\langle \zeta_{\mathbf{k_1}} \zeta_{\mathbf{k_2}} \rangle' = P_\zeta(k_1) \tag{2.49}$$

Moving beyond the power spectrum, in this chapter we will be concerned with the first two non-Gaussian contributions $N = 3$ (bispectrum) and $N = 4$ (trispectrum):

$$\begin{aligned}
\langle \zeta_{\mathbf{k_1}} \zeta_{\mathbf{k_2}} \zeta_{\mathbf{k_3}} \rangle' &= B_\zeta(k_1, k_2, k_3) \\
\langle \zeta_{\mathbf{k_1}} \zeta_{\mathbf{k_2}} \zeta_{\mathbf{k_3}} \zeta_{\mathbf{k_4}} \rangle' &= T_\zeta(k_1, k_2, k_3, s, t)
\end{aligned} \tag{2.50}$$

where s and t are the Mandelstam-like variables as presented at the end of the introduction section. We will now discuss some common shapes of non-Gaussianity that arise in inflationary model building, first for the bispectrum, then the trispectrum.

Primordial bispectra

The simplest way to generate non-Gaussianity is by expanding fluctuations locally in terms of Gaussian fields as follows:

$$\zeta = \zeta_g + \frac{3}{5} f_{\mathrm{NL}}^{\mathrm{loc}} \zeta_g^2 \tag{2.51}$$

where the subscript g denotes Gaussian fields, that have vanishing $(N > 2)$-point functions. Such an expansion gives rise to the so-called local shape of the primordial bispectrum:

$$B^{\mathrm{loc}}(k_1, k_2, k_3) = \frac{6}{5} f_{\mathrm{NL}}^{\mathrm{loc}} \left(P_\zeta(k_1) P_\zeta(k_2) + P_\zeta(k_2) P_\zeta(k_3) + P_\zeta(k_1) P_\zeta(k_3) \right) \tag{2.52}$$

which peaks in the squeezed triangle configuration (e.g. $k_1 \ll k_2 \approx k_3$). Local non-Gaussianity naturally arises in multi-field models of inflation, where extra light fields are present that modulate the dynamics. When the inflaton is able to interact with itself, bispectra of the equilateral type are generated. Their shape is well described by the template:

$$\begin{aligned}
B^{\mathrm{equil}}(k_1, k_2, k_3) = \frac{18}{5} f_{\mathrm{NL}}^{\mathrm{equil}} \Big[\quad &- \left(P_\zeta(k_1) P_\zeta(k_2) + 2 \text{ perms.} \right) \\
&- 2 P_\zeta^{2/3}(k_1) P_\zeta^{2/3}(k_2) P_\zeta^{2/3}(k_3) \\
&+ \left(P_\zeta(k_1) P_\zeta(k_2)^{1/3} P_\zeta(k_3)^{2/3} + 5 \text{ perms.} \right) \Big]
\end{aligned} \tag{2.53}$$

---

[7]From here on out we will use the primed correlator notation to remove the usual factor $(2\pi)^3 \delta_D(...)$.

and peaks in the equilateral momentum configuration ($k_1 \approx k_2 \approx k_3$). All shapes that are due to self-interactions can be systematically classified using the Effective Field Theory of Inflation [23]. Finally, a third shape that is often encountered is the orthogonal shape

$$
\begin{aligned}
B^{\mathrm{ortho}}(k_1, k_2, k_3) = \frac{18}{5} f_{\mathrm{NL}}^{\mathrm{ortho}} \Big[\quad & - \quad 3 \left(P_\zeta(k_1) P_\zeta(k_2) + 2 \text{ perms.} \right) \\
& - \quad 8 P_\zeta^{2/3}(k_1) P_\zeta^{2/3}(k_2) P_\zeta^{2/3}(k_3) \\
& + \quad 3 \left(P_\zeta(k_1) P_\zeta(k_2)^{1/3} P_\zeta(k_3)^{2/3} + 5 \text{ perms.} \right) \Big]
\end{aligned}
$$

(2.54)

which is generated by more exotic models such as Galileon inflation. These three primordial shapes are the most commonly encountered bispectra in inflationary model building, hence in this chapter we will often refer to them together as the common primordial bispectra.

One of the most exciting prospects of measuring non-Gaussianity is the possibility of probing the particle spectrum during inflation. Heavy particles can naturally be present during this high energy phase, leaving a characteristic oscillatory shape in the squeezed limit of the primordial bispectrum, with a frequency set by its mass [7, 9]. More concretely, under exchange of a scalar particle with mass $m^2 > 3H/2$ this oscillation is given by [7]:

$$
\lim_{\mathbf{k}_3 \to 0} \langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle' \propto \quad \frac{1}{k_1^3 k_2^3} \frac{\pi^2}{\cosh^2 \pi\mu} \left(\frac{k_3}{k_1} \right)^{\frac{3}{2}} \times
$$
$$
\left[\left(\frac{k_3}{4k_1} \right)^{-i\mu} \frac{(1 - i \sinh \pi\mu)(\frac{5}{2} - i\mu)(\frac{3}{2} - i\mu)\Gamma(i\mu)}{\Gamma(\frac{1}{2} + i\mu)} + c.c \right].
$$

(2.55)

where $\mu = \sqrt{m^2/H^2 - 9/4}$,[8] clearly showing an oscillation in $\log(k_3/k_1)$ with amplitude, frequency, and phase set by the mass of the scalar particle. The oscillations induced by these heavy fields could be used to follow the evolution of the scale factor, possibly providing additional information about inflation. For this reason, these fields are often referred to as *primordial standard clocks* [24–27]. In [12] a template is

---

[8]In order to stick to conventions in literature we use μ for both the line-of-sight angle $\mu(\mathbf{k})$ and the mass parameter $\mu(m)$. This is unfortunate but their meaning should be sufficiently clear from the context.

proposed for capturing the oscillatory behavior of the bispectrum:

$$B^{\text{clock}}(k_1, k_2, k_3) = \frac{3^{9/2}}{10} \frac{f_{\text{NL}}^{\text{clock}} A_\zeta^2}{(k_1 k_2 k_3)^2} \alpha_{123}^{-1/2} \sin\left(\mu \log\left(\frac{\alpha_{123}}{2}\right) + \delta\right) \Theta(\alpha_{123} - \alpha_0)$$
$$+ 2 \text{ perms.}$$
(2.56)

where $\alpha_{123} = (k_1 + k_2)/k_3$ and the Heaviside step function Θ is included since the oscillation is only present for $\alpha_{123} > 2$ (or permutations). The authors consider a value of $\alpha_0 = 10$ in order to cut off near-equilateral configurations such that overlap with the equilateral shape is reduced. Such a cutoff also drastically reduces the amount of triangle configurations available to measure the signal, thereby lowering its signal-to-noise ratio (SNR). As we will see later in this chapter, the reduced overlap with the equilateral shape does not make up for the loss of SNR and a cutoff $\alpha_0 > 2$ is found to be optimal.

If the exchange particle has a mass $0 \leq m^2 \leq 3H/2$ the oscillatory behavior turns into a power-law [28], called the intermediate shape. Although there exists no analytical expression, the shape is captured well by the following template [12, 28]:

$$B^{\text{int}}(k_1, k_2, k_3) = \frac{6}{5} A_\zeta^2 f_{\text{NL}}^{\text{int}} 3^{\frac{7}{2} - 3\nu} \frac{k_1^2 + k_2^2 + k_3^2}{(k_1 + k_2 + k_3)^{\frac{7}{2} - 3\nu}} (k_1 k_2 k_3)^{-\frac{3}{2} - \nu} \tilde{\Theta}(k_1, k_2, k_3, \alpha_0)$$
(2.57)

where $\nu = \sqrt{9/4 - (m/H)^2} = -i\mu$ and $0 < v < 3/2$. This template interpolates between the behavior of the local ($\nu > 3/2$) and equilateral ($\nu < 1/2$) shapes in the squeezed limit, hence coining the name intermediate. In order to reduce overlap with equilateral shapes, we introduce a cutoff similar to the clock template:

$$\tilde{\Theta}(k_1, k_2, k_3, \alpha_0) = \Theta(\alpha_{123} - \alpha_0) + \Theta(\alpha_{132} - \alpha_0) + \Theta(\alpha_{231} - \alpha_0)$$
(2.58)

In [12] the cutoff was again set to $\alpha_0 = 10$ whereas in this chapter we will find $\alpha_0 = 2$ to maximize the SNR when marginalizing over the other primordial shapes.

Finally, a word about the normalization of the primordial bispectra. The common primordial shapes and intermediate shapes are normalized such that $B(k, k, k) = \frac{18}{5} P_\zeta(k)^2$ in the equilateral configuration. The clock template is normalized such that its amplitude matches that of the intermediate shape for the transition mass $\nu = \mu = 0$ in the squeezed limit.

Primordial trispectra

The next order of correlation functions known as the trispectrum, correlate four fluctuations. As for the bispectrum, the simplest way of generating such a non-Gaussianity is by expanding the primordial field in terms of Gaussian fields to third order:

$$\zeta = \zeta_g + \frac{3}{5} f_{\mathrm{NL}}^{\mathrm{loc}} \zeta_g^2 + \frac{9}{25} g_{\mathrm{NL}}^{\mathrm{loc}} \zeta_g^3 \tag{2.59}$$

in Fourier space, this expansion leads to two distinct local trispectra:

$$T_\zeta^{\tau_{\mathrm{NL}}^{\mathrm{loc}}} = \tau_{\mathrm{NL}}^{\mathrm{loc}} \left[P_\zeta(k_1) P_\zeta(k_3) P_\zeta(s) + P_\zeta(k_1) P_\zeta(k_3) P_\zeta(t) \right] + 6 \text{ perms.} \tag{2.60}$$

$$T_\zeta^{g_{\mathrm{NL}}^{\mathrm{loc}}} = \frac{54}{24} g_{\mathrm{NL}}^{\mathrm{loc}} P_\zeta(k_1) P_\zeta(k_2) P_\zeta(k_3) + 3 \text{ perms.} \tag{2.61}$$

where $s = |\mathbf{s}| = |\mathbf{k}_1 + \mathbf{k}_2|$, $t = |\mathbf{t}| = |\mathbf{k}_1 + \mathbf{k}_4|$ and $u = |\mathbf{u}| = |\mathbf{k}_1 + \mathbf{k}_3|$ are diagonal momenta. For single-field models, the $\tau_{\mathrm{NL}}^{\mathrm{loc}}$ is always generated in case of a nonzero local bispectrum, such that $\tau_{\mathrm{NL}}^{\mathrm{loc}} \geq (\frac{6}{5} f_{\mathrm{NL}}^{\mathrm{loc}})^2$, known as the Suyama-Yamaguchi bound [29]. Furthermore, this shape peaks in the collapsed limit, where one of the diagonals (e.g. s) becomes much smaller than the external momenta. The $g_{\mathrm{NL}}^{\mathrm{loc}}$ amplitude is an independent variable and is characterized by a peak in the double squeezed limit, in which two external momenta become small.

Quartic self-interactions of the inflaton field give rise to equilateral shapes, similar to the equilateral bispectrum. Such interactions can be studied systematically using the EFT of inflation [23]. This effective field theory approach allows one to study the interactions of the Goldstone of time translations $\pi = -\zeta/H$ in the decoupling limit (i.e. without gravity). Typically, quartic interactions for π will imply the existence of cubic interactions that give rise to bispectra that have a much stronger signal than the corresponding trispectrum, unless some contrived mechanism is being considered [30]. Hence, such interactions will likely be detected already by observing the bispectrum. In the presence of an additional light scalar σ, we instead study the EFT of multifield inflation [31]. In this case, it is possible to write down quartic interactions for σ that generate a trispectrum for ζ without generating a larger bispectrum, making the trispectrum the leading order observable non-Gaussianity in such theories [32]. At lowest order in derivatives one finds three interactions $\dot{\sigma}^4$, $\dot{\sigma}^2(\partial\sigma)^2$ and $(\partial\sigma)^4$, which give rise to three distinct trispectra

$$T_\zeta^{\dot\sigma^4} = \frac{221184}{25} g_{NL}^{\dot\sigma^4} \frac{1}{k_1 k_2 k_3 k_4 k_t^5}, \tag{2.62}$$

$$T_\zeta^{\dot\sigma^2(\partial\sigma)^2} = -\frac{27648}{325} g_{NL}^{\dot\sigma^2(\partial\sigma)^2} \frac{k_t^2 + 3(k_3 + k_4)k_t + 12k_3 k_4}{k_1 k_2 (k_3 k_4)^3 k_t^5} (\mathbf{k}_3 \cdot \mathbf{k}_4)$$
$$+ 5 \text{ perms}, \tag{2.63}$$

$$T_\zeta^{(\partial\sigma)^4} = \frac{165888}{2575} g_{NL}^{(\partial\sigma)^4} \frac{2k_t^4 - 2k_t^2 \sum_i k_i^2 + k_t \sum_i k_i^3 + 12k_1 k_2 k_3 k_4}{(k_1 k_2 k_3 k_4)^3 k_t^5} \times$$
$$((\mathbf{k}_1 \cdot \mathbf{k}_2)(\mathbf{k}_3 \cdot \mathbf{k}_4) + 2 \text{ perms}),$$
$$\tag{2.64}$$

where $k_t = k_1 + k_2 + k_3 + k_4$. The amplitudes are normalized such that they match the amplitude of the local shape for $g_{NL} = 1$ in the tetrahedral configuration where $k_i = k$ and $\mathbf{k}_i \cdot \mathbf{k}_j = -k^2/3$: $\langle \zeta^4 \rangle' = \frac{216}{25} P_\zeta(k)^3$ [30]. As we will see, these shapes have a significant overlap. Therefore, when searching for an equilateral trispectrum, it suffices to start off with just one shape.

In [7] it was noted that the oscillatory signature of additional heavy particles is also present in the collapsed limit of the trispectrum. For example, the exchange of a heavy scalar gives rise to a signal:

$$\lim_{s \to 0} \langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \zeta_{\mathbf{k}_4} \rangle' = \frac{1}{4k_1^3 k_2^3 k_3^3 k_4^3} \frac{1}{128\pi} (k_{12} k_{34})^{3/2} \times$$
$$\left[\left(\frac{s^2}{4k_{12}k_{34}} \right)^{i\mu} (1 + i \sinh \pi\mu)(\frac{3}{2} + i\mu)^2 (\frac{5}{2} + i\mu)^2 \right.$$
$$\left. \times \Gamma(-i\mu)^2 \Gamma(\frac{1}{2} + i\mu)^2 + c.c \right]$$
$$\tag{2.65}$$

where $k_{12} = k_1 + k_2$ (note that $s \neq k_{12}$). Again the frequency, amplitude, and phase of the oscillation, are set by the mass through μ. In order to extract such an oscillation from data, we propose the template:

$$T_\zeta^{clock} = \frac{3^{\frac{3}{2}}}{16} g_{NL}^{clock} \frac{1}{k_1^3 k_2^3 k_3^3 k_4^3} (k_{12}k_{34})^{\frac{3}{2}} \Theta(\alpha_{12} - \alpha_0)\Theta(\alpha_{34} - \alpha_0) \sin(\mu \ln(\alpha_{12,34}) + \delta)$$
$$+ 2 \text{ perms}.$$
$$\tag{2.66}$$

where $\alpha_{12,34} = \frac{s^2}{k_{12}k_{34}}$ and $\alpha_{12} = \frac{k_{12}}{s}$. For scalars with an intermediate-mass, the oscillatory behavior in the collapsed limit again turns into a power law as found in [33]:

$$\lim_{s \to 0} \langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \zeta_{\mathbf{k}_4} \rangle' = 4 g_{\mathrm{NL}}^{\mathrm{int}} P_\zeta(k_1) P_\zeta(k_2) P_\zeta(s) \left(\frac{s}{\sqrt{k_1 k_3}} \right)^{3-2\nu} \tag{2.67}$$

To constrain this type of non-Gaussianity we propose the use of a template:

$$T_\zeta^{\mathrm{int}} = 4 \, 3^{\frac{3}{2}-\nu} \, g_{\mathrm{NL}}^{\mathrm{int}} \frac{1}{(k_1 k_2 k_3 k_4)^{\frac{3}{2}}} \frac{1}{s^3} \alpha_{12,34}^{\frac{3}{2}-\nu} \Theta(\alpha_{12} - \alpha_0) \Theta(\alpha_{34} - \alpha_0) + 2 \, \text{perms.} \tag{2.68}$$

As for the bispectrum, we included the Heaviside step functions in order to restrict to the collapsed limit. We will find $\alpha_0 = 2$ to be the cutoff value that optimizes the signal-to-noise ratio for these templates. Furthermore, the intermediate template is normalized to match the $\tau_{\mathrm{NL}}^{\mathrm{loc}}$ shape in the tetrahedral configuration and the amplitude of the oscillatory shape in the collapsed limit is normalized to match the intermediate shape for the mass limit $\nu = \mu = 0$.

2.3.2 Secondary non-Gaussianity

As mentioned before, by the time we enter the Dark Ages non-linear evolution of the density field will generate non-Gaussianity even if we start out with a perfectly Gaussian field. Furthermore, the fact that 21-cm temperature fluctuations depend non-linearly on the underlying fields δ_b, δ_v and δ_{T} causes additional non-Gaussian imprints in the 21-cm signal. Since we are primarily interested in extracting (possibly very weak) primordial non-Gaussianity from this signal, such secondary non-Gaussianity can be considered to be a confusion signal that should be accurately modeled and subtracted in order to adequately determine the primordial contribution. The magnitude and impact of these secondary non-Gaussianities were previously studied for the bispectrum in [11, 14]. There, it was found that the secondary contribution is several orders of magnitude stronger than the primordial signal. Including the amplitude of the secondary bispectrum as nuisance parameters and marginalizing over them reduces the signal-to-noise ratio of the primordial amplitude. In this chapter, we improve upon some of the assumptions made in those previous works, by including the effects of electron fraction fluctuations and baryonic pressure. The former has been studied in section 2.2.2 and its effect is captured in the coefficients $\mathcal{C}_{i,j}^{\mathrm{T}}$. The latter will be studied next. After having set up the framework for higher-order perturbation theory with baryonic pressure, we will be able to derive both the secondary bispectrum as well as for the first time the secondary trispectrum of 21-cm temperature fluctuations during the Dark Ages.

Perturbation theory including baryonic pressure

By means of equation (2.45), we have a direct perturbative relation between 21-cm fluctuations and baryon density contrast and velocity divergence up to third order. However, we have so far been agnostic about the functional form of $\delta_b^{(n)}$ and $\theta_b^{(n)}$. In earlier work [11, 14], two assumptions are made about the functional form. First, it assumed that the time-dependence of the baryonic fluctuations is identical to that of CDM fluctuations so that the n-th order perturbations can be written as [34]:

$$\delta_b^{(n)}(z, \mathbf{k}) = a^n(z)\delta_{n,b}(\mathbf{k}), \qquad \theta_b^{(n)}(z, \mathbf{k}) = \mathcal{H}(z)a^n(z)\theta_{n,b}(\mathbf{k}). \qquad (2.69)$$

Note that the functional form is product separable in temporal and spatial (momentum) dependence. The latter is contained in δ_n and θ_n, which in standard perturbation theory read:

$$\delta_{n,b}(\mathbf{k}) = \int_{\mathbf{q}_1 \cdots \mathbf{q}_n} (2\pi)^3 \delta_{\mathrm{D}}(\mathbf{k} - \mathbf{q}_{1\cdots n}) F_{n,b}(\mathbf{q}_1, \ldots, \mathbf{q}_n)\delta_{1,b}(\mathbf{q}_1) \cdots \delta_{1,b}(\mathbf{q}_n) \qquad (2.70)$$

$$\theta_{n,b}(\mathbf{k}) = \int_{\mathbf{q}_1 \cdots \mathbf{q}_n} (2\pi)^3 \delta_{\mathrm{D}}(\mathbf{k} - \mathbf{q}_{1\cdots n}) G_{n,b}(\mathbf{q}_1, \ldots, \mathbf{q}_n)\delta_{1,b}(\mathbf{q}_1) \cdots \delta_{1,b}(\mathbf{q}_n), \qquad (2.71)$$

where $\mathbf{q}_{1\cdots n} \equiv \mathbf{q}_1 + \cdots + \mathbf{q}_n$ and the *baryonic* momentum kernels $F_{n,b}$ and $G_{n,b}$ are symmetric under interchanging the momenta. The second assumption made in previous works is that the baryonic kernels are equivalent to that of CDM so that they can be replaced by the latter: $F_{n,b} \to F_{n,c}$ and $G_{n,b} \to G_{n,c}$, which can trivially be obtained from a simple recursion relation [34].

However, in reality, baryons are very different from CDM in their physical nature. Where CDM is pressureless, baryons constitute pressure. In particular, on scales comparable to the so-called baryonic Jeans scale λ_{J}, the pressure becomes considerable and suppresses the growth of fluctuations due to gravity. On scales smaller than the Jeans scale, pressure competes with gravity to generate acoustic sound waves and the notion of growing density fluctuations becomes ill-defined.

To take the first step in the direction of accounting for pressure effects, we use the formalism for baryonic fluctuations constructed in [35].[9] Within this formalism, the Jeans scale is assumed to be constant, i.e. $\lambda_{\mathrm{J}} \neq \lambda_{\mathrm{J}}(z)$. Although strictly speaking this assumption becomes incorrect once the gas temperature decouples from the photon temperature, it has the convenient consequence that the time-dependence of baryonic fluctuations is still equivalent to that of CDM, $\delta_b^{(n)} \propto a^n$, and the results of section 2.2.3 are still applicable. Then, the baryonic fields δ_b and θ_b can be expanded

---

[9]We will only highlight the required essentials of the formalism, for details, we refer to [35].

as:

$$\delta_b(z,\mathbf{k}) = \sum_{n=1}^{\infty} a^n(z) g_n(\mathbf{k}) \delta_{n,c}(\mathbf{k}), \qquad \theta_b(z,\mathbf{k}) = \sum_{n=1}^{\infty} \mathcal{H}(z) a^n(z) h_n(\mathbf{k}) \theta_{n,c}(\mathbf{k}),$$

(2.72)

where we have defined the *filtering* functions:[10]

$$g_n(\mathbf{k}) = \frac{\delta_{n,b}(\mathbf{k})}{\delta_{n,c}(\mathbf{k})}, \qquad h_n(\mathbf{k}) = \frac{\theta_{n,b}(\mathbf{k})}{\theta_{n,c}(\mathbf{k})}.$$

(2.73)

Note that while we solely require g_1 at first order (as the linear density contrast and velocity divergence are trivially related as $\theta_b = -\mathcal{H}\delta_b$), at higher order we require different filtering functions for δ_b (i.e. g_n) and θ_b (h_n).

We expect the filtering functions to be strongly decreasing for scales around and beyond the Jeans scale ($k/k_\mathrm{J} \gg 1$), as pressure suppresses the evolution of baryonic density fluctuations relative to those in CDM on these scales. In addition, far away from the Jeans scale ($k/k_\mathrm{J} \ll 1$), we assume the behavior of baryons and CDM to be equivalent, i.e. $g_n, h_n \to 1$. At first order, the filtering function is given by:

$$g_1(k) = \frac{1}{1 + k^2/k_\mathrm{J}^2},$$

(2.74)

which indeed contains the limiting behavior described above. At second order, the function g_2 can be written in terms of g_1 as:

$$g_2(\mathbf{k}) = \sigma(k) \left[1 + \frac{7}{3} \frac{\delta'_{2,c}(\mathbf{k})}{\delta_{2,c}(\mathbf{k})} \right], \qquad \sigma(k) \equiv \frac{1}{10/3 + (k/k_\mathrm{J})^2},$$

(2.75)

where $\delta'_{2,c}$ is given by (the prime does not denote a time derivative):

$$\delta'_{2,c}(\mathbf{k}) \equiv \int_{\mathbf{q}_1 \mathbf{q}_2} (2\pi)^3 \delta_\mathrm{D}(\mathbf{k} - \mathbf{q}_{12}) \, \mathcal{F}_2^{(s)}(\mathbf{q}_1, \mathbf{q}_2) \, \delta_{1,c}(\mathbf{q}_1) \delta_{1,c}(\mathbf{q}_2).$$

(2.76)

The modified kernel $\mathcal{F}_2^{(s)}(\mathbf{q}_1, \mathbf{q}_2)$ (where the superscript (s) denote the symmetrization of the kernel) reads:

$$\mathcal{F}_2^{(s)}(\mathbf{q}_1, \mathbf{q}_2) \equiv \left[F_2^{(s)}(\mathbf{q}_1, \mathbf{q}_2) + \frac{3}{14} \frac{k^2}{k_\mathrm{J}^2} \right] g_1(\mathbf{q}_1) g_1(\mathbf{q}_2),$$

(2.77)

---

[10]In principle, these filtering functions depend on time. Therefore, they are often defined as: $g_n(z,\mathbf{k}) = \delta_b^{(n)}(z,\mathbf{k})/\delta_c^{(n)}(z,\mathbf{k})$, including time dependence. However, as shown in [35], when decaying modes are omitted and the Jeans scale is taken to be constant, the filtering functions become time-independent as well, justifying this definition.

which reduces to the CDM kernel $F_2^{(s)}(\mathbf{q}_1, \mathbf{q}_2)$ in the limit $k/k_J \ll 1$. The function $h_2(\mathbf{k})$ can be written as:

$$h_2(\mathbf{k}) = \frac{1}{\theta_{2,c}(\mathbf{k})} \left[\int_{\mathbf{q}_1 \mathbf{q}_2} (2\pi)^3 \delta_D(\mathbf{k} - \mathbf{q}_{12}) \Big[1 + \vartheta(\mathbf{q}_1, \mathbf{q}_2) \Big] \delta_{1,b}(\mathbf{q}_1) \delta_{1,b}(\mathbf{q}_2) \right] - 2 \frac{\delta_{2,b}(\mathbf{k})}{\theta_{2,c}(\mathbf{k})},$$
(2.78)

where we have defined $\vartheta(\mathbf{q}_1, \mathbf{q}_2) \equiv (\mathbf{q}_1 \cdot \mathbf{q}_2)(q_1^2 + q_2^2)/2q_1^2 q_2^2$. In the limit $k/k_J \to 0$, we obtain $h_2 \to 1$ as expected.

The higher order perturbations $\delta_b^{(3)}, \theta_b^{(3)}$ that show up in the secondary trispectrum will also be modified as compared to the CDM expressions.[11] However, as we will see later when comparing forecasts of the bispectrum with and without baryonic pressure effects, the differences are small. Hence, for computational simplicity, we will not include baryonic pressure in our forecasts for the trispectrum. With the above perturbation expansions, we are now able to write down the secondary contributions to the non-Gaussianity of the 21-cm signal.

Secondary bispectrum

The lowest order non-Gaussian statistic of 21-cm brightness temperature fluctuations, i.e. the 21-cm bispectrum $B_{\delta T}$, now consists of a primary, primordial contribution:

$$\langle \delta T_{21}^{(1)}(\mathbf{k}_1) \delta T_{21}^{(1)}(\mathbf{k}_2) \delta T_{21}^{(1)}(\mathbf{k}_3) \rangle' = \left(\prod_{i=1}^{3} c_1^{(1)}(\mathbf{k}_i) \mathcal{M}_b(\mathbf{k}_i) \right) \times B_\zeta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$$
(2.79)

where $\mathcal{M}_b(\mathbf{k})$ is the linear transfer function for baryonic fluctuations. Moreover, there is also a higher-order, secondary contribution:

$$\begin{aligned}
\langle \delta T_{21}(\mathbf{k}_1) \delta T_{21}(\mathbf{k}_2) \delta T_{21}(\mathbf{k}_3) \rangle'_{\text{sec.}} &= \langle \delta T_{21}^{(1)}(\mathbf{k}_1) \delta T_{21}^{(1)}(\mathbf{k}_2) \delta T_{21}^{(2)}(\mathbf{k}_3) \rangle' + 2 \text{ perms.} \\
&= c_1^{(1)}(\mathbf{k}_1) c_1^{(1)}(\mathbf{k}_2) c_1^{(2)}(\mathbf{k}_3) \times 2\tilde{\mathcal{F}}_2^{(s)}(\mathbf{k}_1, \mathbf{k}_2) P_c(\mathbf{k}_1) P_c(\mathbf{k}_2) \\
&\quad + c_1^{(1)}(\mathbf{k}_1) c_1^{(1)}(\mathbf{k}_2) c_2^{(2)}(\mathbf{k}_3) \times 2\tilde{\mathcal{G}}_2^{(s)}(\mathbf{k}_1, \mathbf{k}_2) P_c(\mathbf{k}_1) P_c(\mathbf{k}_2) \\
&\quad + c_1^{(1)}(\mathbf{k}_1) c_1^{(1)}(\mathbf{k}_2) \left[c_3^{(2)}(\mathbf{k}_3, -\mathbf{k}_1) + c_3^{(2)}(\mathbf{k}_3, -\mathbf{k}_2) \right] g_1^2(k_1) g_1^2(k_2) P_c(\mathbf{k}_1) P_c(\mathbf{k}_2) \\
&\quad + 2 \text{ perms.}
\end{aligned}$$
(2.80)

---

[11]See Appendix B of [35], Eqs. B23 and B25.

We have defined the modified kernels $\tilde{\mathcal{F}}_2^{(s)}$ and $\tilde{\mathcal{G}}_2^{(s)}$ as:

$$\tilde{\mathcal{F}}_2^{(s)}(\mathbf{k}_1, \mathbf{k}_2) \equiv g_1(k_1)g_1(k_2)\sigma_2(|\mathbf{k}_{12}|) \left[F_2^{(s)}(\mathbf{k}_1, \mathbf{k}_2) + \frac{7}{3}\mathcal{F}_2^{(s)}(\mathbf{k}_1, \mathbf{k}_2) \right],$$

$$\tilde{\mathcal{G}}_2^{(s)}(\mathbf{k}_1, \mathbf{k}_2) \equiv 2g_1(k_1)g_1(k_2)\sigma_2(|\mathbf{k}_{12}|) \left[F_2^{(s)}(\mathbf{k}_1, \mathbf{k}_2) + \frac{7}{3}\mathcal{F}_2^{(s)}(\mathbf{k}_1, \mathbf{k}_2) \right]$$
$$- g_1^2(k_1)g_1^2(k_2)[1 + \vartheta(\mathbf{k}_1, \mathbf{k}_2)], \tag{2.81}$$

In the limit $\lambda_J \to \infty$ and $P_c \to P_b$ this restores to the previously derived secondary bispectrum of 21-cm brightness temperature fluctuations [11, 14].

Secondary trispectrum

The secondary trispectrum is composed of 11 different contributions, which can be summarized schematically as:

$$\langle \delta T_{21}(\mathbf{k}_1)\delta T_{21}(\mathbf{k}_2)\delta T_{21}(\mathbf{k}_3)\delta T_{21}(\mathbf{k}_4)\rangle'_{\text{sec.}} = T_{\delta_1\delta_1\delta_2\delta_2} + T_{\delta_1\delta_1\theta_2\theta_2} + T_{\delta_1\delta_1\delta_2\theta_2}$$
$$+ T_{\delta_1\delta_1\delta_2[\delta_1]^2} + T_{\delta_1\delta_1\theta_2[\delta_1]^2} + T_{\delta_1\delta_1[\delta_1]^2[\delta_1]^2}$$
$$+ T_{\delta_1\delta_1\delta_1[\delta_1]^3} + T_{\delta_1\delta_1\delta_1[\delta_1\delta_2]} + T_{\delta_1\delta_1\delta_1[\delta_1\theta_2]}$$
$$+ T_{\delta_1\delta_1\delta_1\delta_3} + T_{\delta_1\delta_1\delta_1\theta_3}, \tag{2.82}$$

where for example

$$T_{\delta_1\delta_1\delta_2\theta_2} = f(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4)\langle \delta_b^{(1)}(\mathbf{k}_1)\delta_b^{(1)}(\mathbf{k}_2)\delta_b^{(2)}(\mathbf{k}_3)\theta_b^{(2)}(\mathbf{k}_4)\rangle + 11 \text{ perms.} \tag{2.83}$$

Here the function f is composed of the coefficients $c_i^{(j)}$ introduced above. The explicit expressions for all 11 contributions can be found in Appendix 2.D.

2.4 The information content of the Dark Ages

In this section we will determine the amount of information in 21-cm brightness temperature fluctuations during the Dark Ages available to constrain primordial non-Gaussianity.[12] First, we will determine the total amount of information available when neglecting any form of signal confusion (either instrumental, systematic, or secondary), giving an ultimate upper bound for the sensitivities that can be achieved. Subsequently, using our improved model of the 21-cm signal, we determine the amount of information when including signal confusion due to secondary non-Gaussianities. Finally, we will determine the sensitivity of different experimental scenarios.

---

[12]The code used to perform the numerical calculations is openly available on GitHub via this link.

2.4.1   Total primordial information content

In order to determine the total information content, we will assume a cosmic variance limited experiment of the Dark Ages between $30 \leq z \leq 100$ and take the largest scale to be $k_{\min} = 0.01$ Mpc^{-1}, since on larger scales one has to include fluctuations in the CMB temperature and non-relativistic perturbation theory breaks down.[13] When including the secondary bispectrum and trispectrum, following [36] we consider 14 redshift bins (i.e. $\Delta z = 5$), assuming that they are sufficiently uncorrelated such that the covariance matrix can be taken to be diagonal in redshift space. For more details about the Fisher analysis and how to evaluate the Fisher matrix for both the bispectrum and trispectrum, we refer to Appendix 2.C.

The results for the common primordial bispectra are summarized in Table 2.2. The sensitivity to the amplitude of the primordial bispectrum is given on the diagonal. The increased sensitivity to the local shape as compared to the other two common shapes, can be understood as due to a slightly enhanced scaling with $k_{\max}$ [15]. Furthermore, on the off-diagonal, we summarize the overlap coefficient between the common primordial bispectra.

	f_{NL}^{loc}	f_{NL}^{equil}	f_{NL}^{ortho}
f_{NL}^{loc}	4.22×10^{-6}	0.29	-0.33
f_{NL}^{equil}		2.88×10^{-5}	0.19
f_{NL}^{ortho}			2.10×10^{-5}

Table 2.2: Diagonal*: sensitivity to the amplitude of the common primordial shapes.* Off-diagonal*: cosine* (2.111) *between common primordial shapes. We assume a cosmic variance limited experiment between* $30 \leq z \leq 100$ *and* $0.01 \leq k \leq 300$ *Mpc*$^{-1}$.

We now consider the massive scalar exchange template for $m \geq 3H/2$ (i.e. equation (2.56)). As mentioned, the authors of [12] suggested the use of a cutoff $\alpha_0 = 10$ in order to remove overlap with the equilateral shape equation (2.53). Such a cutoff also drastically reduces the amount of triangles available to observe, hence reducing the signal-to-noise ratio. Using the Fisher analysis we are able to determine the amount of overlap between the clock template and the common primordial shapes that we aim to look for in future experiments. In Figure 2.8a we present this overlap as a function of μ for both the cutoff $\alpha_0 = 10$ and $\alpha_0 = 2$. We find that although the overlap with the equilateral shape is almost completely removed for the former

---

[13]The survey volume sets the largest scale to be only slightly bigger at $k = 0.005$ Mpc^{-1}. Hence, from an observational perspective, one gains very little from going beyond our choice of $k_{\min}$.

	$\tau_{\mathrm{NL}}^{\mathrm{loc}}$	$g_{\mathrm{NL}}^{\mathrm{loc}}$	$g_{\mathrm{NL}}^{\dot{\sigma}^4}$	$g_{\mathrm{NL}}^{\dot{\sigma}^2(\partial\sigma)^2}$	$g_{\mathrm{NL}}^{(\partial\sigma)^4}$
$\tau_{\mathrm{NL}}^{\mathrm{loc}}$	1.79×10^{-7}	0.00	0.00	0.00	0.00
$g_{\mathrm{NL}}^{\mathrm{loc}}$		0.02	0.13	0.11	0.06
$g_{\mathrm{NL}}^{\dot{\sigma}^4}$			0.229	0.95	0.65
$g_{\mathrm{NL}}^{\dot{\sigma}^2(\partial\sigma)^2}$				0.254	0.82
$g_{\mathrm{NL}}^{(\partial\sigma)^4}$					0.0609

Table 2.3: Diagonal*: sensitivity σ to the amplitude of primordial trispectra. Off-diagonal: cosine (2.111) between common primordial trispectra. We assume a cosmic variance limited experiment between $30 \leq z \leq 100$ and $0.01 \leq k \leq 300 \ Mpc^{-1}$.*

case, the overlap with the other shapes is increased for some mass values compared to the latter. In 2.7a we show the sensitivity to the amplitude of the clock template, which clearly shows that a larger cutoff is favorable only for masses $\mu \lesssim 1$. A similar observation can be made for the intermediate template (Figure 2.8b and 2.7b) where we again find $\alpha_0 = 2$ to optimize the sensitivity except for some small improvement with $\alpha_0 = 10$ when $\nu \gtrsim 1$. Hence, in what follows we will take $\alpha_0 = 2$.

Proceeding to the trispectrum, in Table 2.3 (on the diagonal) we present the ultimate sensitivity of 21-cm temperature fluctuations from the Dark Ages to the amplitude of the local and equilateral trispectra. We find the $\tau_{\mathrm{NL}}^{\mathrm{loc}}$ shape to be much better constrained than the other shapes. This can again be understood as due to an enhanced scaling [15], which we will comment on later in this subsection. The overlap between these primordial trispectra is given by the off diagonals, where we see that the enhanced scaling of the $\tau_{\mathrm{NL}}^{\mathrm{loc}}$ shape makes it completely orthogonal to the other shapes, making it a very clean signal to look for.

The ultimate sensitivities for the massive exchange trispectra are given in Figure 2.9. Although it's clear that the clock trispectrum does not outperform the clock bispectrum, we see that the intermediate template enjoys an enhanced scaling for masses $\nu > 3/4$. Such a scaling can improve the sensitivity to the trispectrum past the sensitivity to the bispectrum. However, for ν close to $3/2$, the shape acquires significant overlap with the $\tau_{\mathrm{NL}}^{\mathrm{loc}}$ shape, which drastically reduces the sensitivity after marginalization making the sensitivity on par with that of the bispectrum. The overlap of the exchange trispectra with the local and equilateral trispectra is summarized in Figure 2.10, which shows how the intermediate template interpolates between the local and equilateral shapes depending on the mass ν.

The results presented in this subsection give an indication of how well 21-cm

brightness temperature fluctuations from the Dark Ages are able to constrain primordial cosmology if we were able to harvest all of its information down to the Jeans scale $k_{\max} = 300\,\mathrm{Mpc}^{-1}$, demonstrating the ability of 21-cm cosmology to achieve the exquisite sensitivity necessary to advance our understanding of the primordial universe, possibly even paving a way towards the Cosmological Collider as noticed previously by [12].

2.4.2   Including secondary non-Guassianity

Although 21-cm brightness temperature fluctuations are still a relatively pristine tracer of the underlying density field sourced by inflation, non-linear evolution such as due to gravitational collapse or velocity effects, will make even perfectly Gaussian fluctuations non-Gaussian over time, and between the surface of last scattering at $z \approx 1100$ and the Dark Ages $z \approx 200$ there is a considerable amount of time for such secondary non-Gaussianities to be generated. When trying to extract primordial non-Gaussianity from 21-cm observations, such secondary effects enter as confusion signals, introducing extra nuisance parameters that should be modeled and accounted for when analyzing data. Secondary non-Gaussianity of the 21-cm signal has already been studied in the context of the bispectrum before [11, 14]. In the previous sections, we have improved upon some of the assumptions that went into those works in order to more accurately model the secondary bispectrum, as well as to extend this to the secondary trispectrum. In this subsection, we will study the impact of secondary non-Gaussianity on the constraining power of the 21-cm signal from the Dark Ages.

The amount of additional nuisance parameters depends on how accurately we are able to model the secondary non-Gaussianity. Most optimally, we might be able to model the internal parameters of secondary non-Gaussianity ($\bar{T}_{21}$ and $\alpha_{i,j}$ in 2.2.4) to such precision that the only free parameter of the secondary N-point function is its amplitude. In that case, the secondary N-point function can be treated as a single shape. More realistically, we obtain the best fit for the internal parameters such that their residuals become the free parameters of the secondary correlation function [14]:

$$F_{\delta T}(\mathbf{k}_1, ..., \mathbf{k}_N) = F_{\delta T}^{\mathrm{sec},0}(\mathbf{k}_1, ..., \mathbf{k}_N) + \sum_i \Delta A_i \frac{\partial F_{\delta T}^{\mathrm{sec}}}{\partial A_i} \qquad (2.84)$$

such that we end up with an individual shape $\frac{\partial F_{\delta T}^{\mathrm{sec}}}{\partial A_i}$ for every internal parameter. Most pessimistically, one could consider every independent shape present in the secondary N-point function. Due to the large number of shapes (already 21 for the bispectrum), calculating the full Fisher matrix quickly becomes intractable for higher

N-point functions. In the following we will adopt the best-fit approach of [14], giving rise to a four-parameter secondary bispectrum and seven-parameter secondary trispectrum. We have improved upon previous work by including free electron fraction perturbations to model the internal parameters of secondary non-Gaussianity, and we include baryonic pressure (in the bispectrum).

We quantify the impact of including secondary non-Gaussianity on the signal-to-noise ratio using the signal-to-noise degradation factor (SND) defined in equation (2.110). It's interesting to note that although the signal-to-noise degradation at each redshift might be relatively large when co-adding the redshifts the degradation is significantly reduced. This can be understood due to the difference in how the secondary shapes scale with z as compared to the primordial shapes: going from equation (2.94) to equation (2.96), we see that the amount of signal from the primordial shape does not depend on z whereas the secondary shapes (including off-diagonal contributions to the Fisher matrix) do. Hence, when adding information from redshift slices, one can also start distinguishing the primordial and secondary shapes by their redshift dependence.

In Figure 2.11 we plot the amount of degradation as a function of k_{max} for the primordial bispectra. We find that including baryonic pressure slightly worsens the signal-to-noise degradation as compared to the simplified approach of [14], but not by much. We expect the same to be true for the trispectrum, so in order to keep the trispectrum calculations more tractable and efficient, we neglect baryonic pressure effects in the upcoming trispectrum analysis. Nevertheless, in an actual data analysis, one should model the bi- and trispectrum as accurately as possible, in order to reliably extract the primordial signal. Furthermore, we find the four parameters of the secondary bispectrum to contribute evenly to the signal-to-noise degradation (meaning their cosine (2.111) with f_{NL} is similar).

We summarize the sensitivity to the amplitude of the common primordial bispectra as a function of k_{max} in Figure 2.4, clearly showing the small loss of signal-to-noise due to the secondary bispectrum. Similar results are presented for the clock and intermediate template in Figure 2.12 and Figure 2.13 respectively. In the top panels of those figures, we see that once overlap with the common primordial shapes is taken into account, the additional effect of the secondary bispectrum (difference between dotted and solid lines in the figures) becomes negligible. We thus conclude that for the exchange templates, the overlap with other primordial shapes is more severe than any overlap with secondary shapes.

We now investigate the impact of the secondary trispectrum on the previously
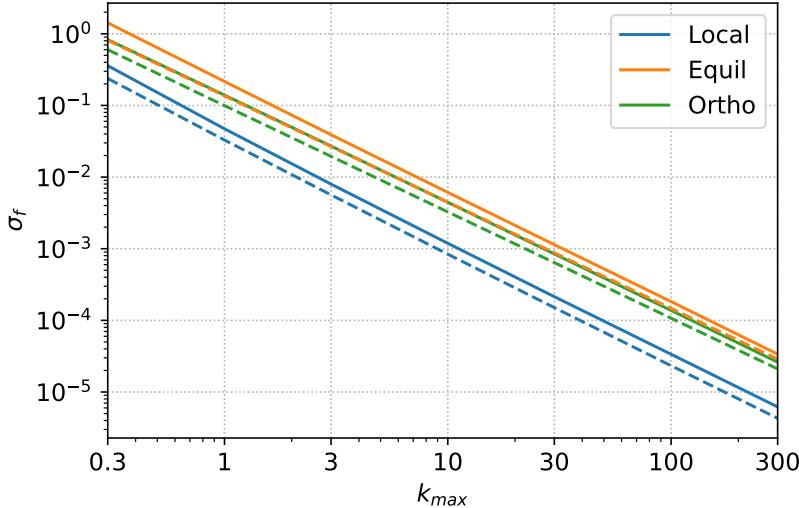
Figure 2.4: *The sensitivity to the amplitude of the common primordial bispectra as a function of k_{max} before (dashed) and after (solid) marginalizing over the 4-parameter secondary bispectrum including baryonic pressure effects. The solid green line coincides with the dashed orange line.*

obtained sensitivities to primordial trispectra. In Figure 2.14 we summarize the signal-to-noise degradation after marginalizing over the seven residual parameters of the secondary trispectrum. Due to the more complex shapes of the trispectra in k-space as compared to the bispectra, we see little overlap with the secondary trispectra, especially for higher values of k_{max}. The small overlap is reflected in the small loss of sensitivity to the amplitude of the common shapes in Figure 2.5. As for the bispectrum, we find similar overlap for all parameters. Results for the clock and intermediate template are summarized in Figure 2.15 and Figure 2.16 respectively. Similar to the case of the bispectrum, in the left figures we see that when including the overlap of the common primordial trispectra, the additional effect of the secondary trispectra is negligible. What is different from the case of the bispectrum is the enhanced scaling for the intermediate template at mass values $\nu > 0.75$.

2.4.3   Forecasting experiments

Until now we have studied the signal-to-noise with which we can measure primordial non-Gaussianity as a function of the maximum accessible momentum k_{max}. In reality, however, there is a difference between the radial (or redshift) component of the momenta along the line-of-sight $k_{\parallel}$ and the perpendicular (or angular) compo-
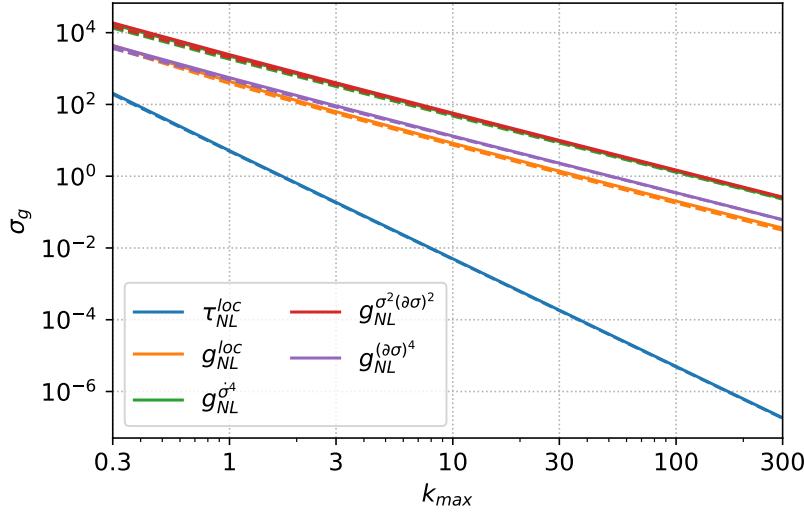
Figure 2.5: *The sensitivity to the amplitude of the common primordial trispectra as a function of $k_{\max}$ before (dashed) and after (solid) marginalizing over the 7-parameter secondary trispectrum. Green and red lines coincide. The enhanced scaling of the τ_{NL}^{loc} shape is clearly visible.*

nent $k_\perp$. The resolution along these components is set by different properties of the experimental setup. The line-of-sight resolution is determined by our ability to distinguish redshift slices, which in practice corresponds to the size of the frequency bins set by the window size $\delta\nu$ [36]:

$$k_{\max}^{\parallel} \approx \sqrt{\frac{17}{3}} \frac{1}{20\delta\nu\sqrt{1+z}} \qquad (2.85)$$

with $\delta\nu$ in units of MHz. Furthermore, the angular resolution R is set by the baseline b (the distance between two receivers in an interferometer) of the experiment through:

$$R = \frac{\lambda_{\text{obs}}}{b} \qquad (2.86)$$

which yields the maximum perpendicular mode

$$k_{\max}^{\perp} \approx \frac{2\pi\nu_0 b}{d(z)(1+z)c} \qquad (2.87)$$

with b in units of km and where $\nu_0 \approx 1420 \times 10^6$ Hz is the frequency of the 21-cm signal, c the speed of light in km/s and $d(z)$ the comoving distance to redshift z in

units of Mpc. In practice, it is much easier to improve on the window $\delta\nu$ than it is to increase the baseline. In Figure 2.6 we show k_{max} as a function of baseline and window size at several redshifts.
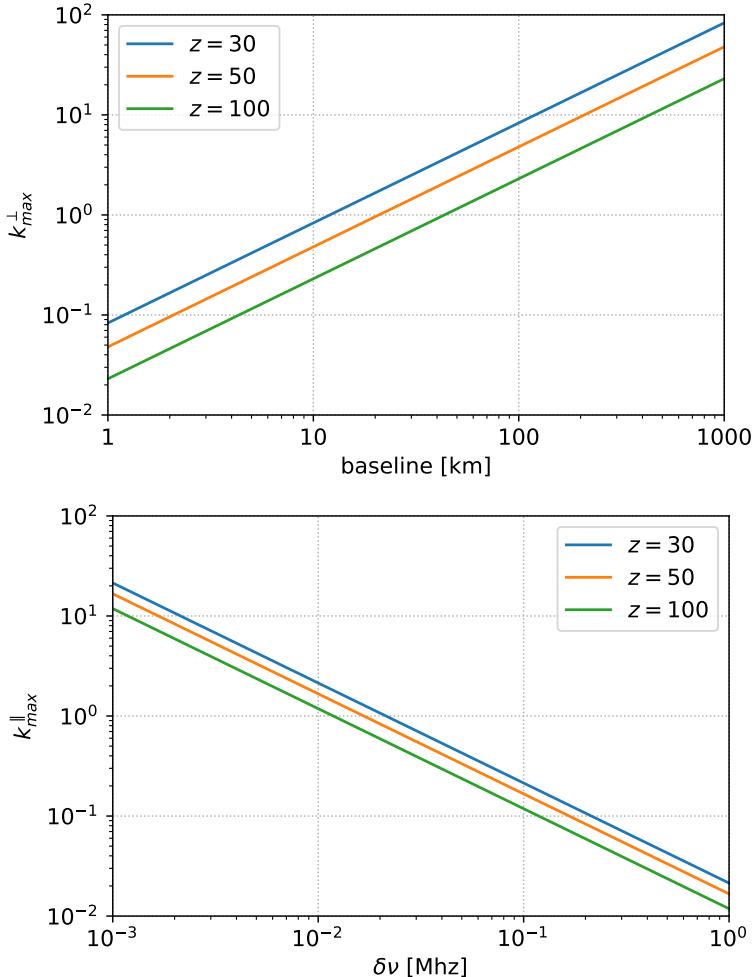
Figure 2.6: *The maximal wavenumber k_{max} that can be probed in a certain direction for several redshifts.* Left: *perpendicular $k_{\mathrm{max}}^{\perp}$ as a function of baseline in kilometers.* Right: *parallel (line-of-sight) $k_{\mathrm{max}}^{\parallel}$ as a function of window size $\delta\nu$.*

Figure 2.17 shows the forecast sensitivity to the amplitude of the three common primordial bispectra. We find that already for a reasonably sized array with

a baseline of several kilometers 21-cm observations from the Dark Ages can improve current as well as future CMB constraints by orders of magnitude. Notably, $\sigma_{f_{\mathrm{NL}}^{\mathrm{loc}}} \sim O(10^{-2})$ can in principle be achieved with a baseline smaller than 10 kilometers, allowing for a test of the famous Maldacena consistency condition.

Figures 2.18 and 2.19 summarize the ability to detect primordial features due to the exchange of massive particles during inflation for different baseline and window size. Depending on the nature of the exchange coupling the size of the amplitude of the non-Gaussian signal can be anywhere between $f_{\mathrm{NL}} \ll 10^{-2}$ and $f_{\mathrm{NL}} > O(1)$ (see section E. of [12] for a clear discussion on the different possibilities). We see that a baseline of $100 - 1000$ km could already provide a clear picture of the particle content of the early universe, opening up an entirely new window into the physics of inflation.

In Figures 2.20 and 2.21 we present for the first time forecasts of the sensitivity of different experimental setups to the amplitude of primordial trispectra, taking into account the secondary trispectrum of 21-cm brightness temperature fluctuations. We find that for most trispectra it proves hard to do better than $\sigma_g \sim O(1)$, a clear exception being the $\tau_{\mathrm{NL}}^{\mathrm{loc}}$ local shape, which exhibits enhanced scaling with k_{max} resulting in sensitivities on par with that of the bispectrum.

Moving on to the trispectrum of massive exchange, we show the results in Figures 2.22 and 2.23. We conclude that the oscillatory clock template can realistically only be constrained up to a sensitivity of $\sigma_g \sim O(10)$, still many orders of magnitude better than CMB observations (likely $\sim O(10^4)$ in the future). However, the intermediate trispectrum of massive exchange can reach sensitivities that are much more similar to the bispectrum sensitivities, allowing one to probe the same masses through both the bispectrum and trispectrum. Although we have not explicitly included spinning particles in this chapter, having two independent probes of the same mass could be used to break the degeneracy between the mass and spin of the particle, which both contribute to the amplitude of the signal [7].

2.5 Conclusions and Outlook

Measurements of primordial non-Gaussianity play a key role in unraveling the mysteries of the early universe. Until now, the most stringent constraints on the amount of pnG come from observations of the cosmic microwave background radiation, which are still consistent with zero. Future CMB experiments (e.g. SO and CMBS4) could improve constraints by an order of magnitude, possibly zeroing in on non-zero

pnG. However, in order to accurately distinguish different non-Gaussian signatures such as the imprint of massive particles present during inflation, sensitivities are required that can simply not be achieved with CMB measurements, due to the limited amount of modes available. To advance our understanding of the primordial universe, we ought to employ additional probes. Arguably the most promising avenue to this end is the use of 21-cm brightness temperature fluctuations. During a part of the cosmological Dark Ages (between $30 \leq z \leq 200$), these tiny fluctuations can in principle be observed in absorption to the CMB radiation. The result is a large observable 3D volume of linear modes that trace the primordial initial conditions, which could constrain primordial non-Gaussianity with unprecedented precision, opening up an exciting new window into the physics of the early universe.

In order to extract such primordial information from the tracer field, it's required to accurately model the physics of the hydrogen gas. Furthermore, the non-linear relation between the initial conditions and the 21-cm brightness temperature fluctuations as well as the gravitational collapse of the matter field induce secondary non-Gaussianity, swamping to primordial signal that we aim to observe. In this chapter, we (re)addressed some of these aspects and made several advances. Specifically, we have:

- improved analytical modeling of the hydrogen gas by including the effect of free electron fraction perturbations, which turns out to be sizable (Figure 2.2) and hence should be included in any detailed analysis of 21-cm observations of the Dark Ages

- derived the perturbation expansion of the 21-cm tracer field up to third order in the underlying matter field, while including baryonic pressure effects on scales close to the Jeans scale

- thereby obtained an improved expression for the secondary bispectrum of 21-cm brightness temperature fluctuations during the Dark Ages

- derived, for the first time, the secondary trispectrum of the 21-cm tracer field

With this improved modeling of the tracer field and the newly obtained secondary bi- and trispectrum, we asses the information content of 21-cm fluctuations to constrain primordial physics, under the most optimal conditions. More specifically, we perform a Fisher analysis to determine the error bars on the amplitude of various primordial bi- and trispectra in a potential 21-cm Dark Ages experiment, marginalizing over the parameters of secondary non-Gaussianity. From this analysis, we draw the following conclusions:

- Although secondary non-Gaussianity swamps the primordial signal, the primordial signal can be extracted without a big loss of signal-to-noise both for the bispectrum and the trispectrum

- tomography of 21-cm fluctuations can in principle improve constraints on primordial non-Gaussianity by several orders of magnitude, already for relatively low k_{max}, as noted before in refs. [11, 12, 14]

- the enhanced scaling of some trispectra with k_{max} [15] indeed manifests itself in 21-cm observations.

- our Fisher forecasts demonstrate the importance of marginalizing over primordial shapes as compared to secondary shapes when considering the massive particle exchange templates

Our results clearly show the potential of the 21-cm signal from the Dark Ages to improve constraints on primordial non-Gaussianity and thereby our understanding of the infant universe. Many interesting directions of research remain in order to better understand the intricacies of the 21-cm Dark Ages signal. We would like to reiterate some of the assumptions made in this chapter that can be improved upon as well as some other possible directions for future research:

- we have included baryonic pressure effects to the secondary bispectrum along the lines of [35]. One of the assumptions made there is the constancy of the baryonic Jeans scale k_J. In reality, this scale depends on time (redshift) and a more accurate modeling of secondary non-Gaussianity of the 21-cm signal should include this time-dependence

- Our expression for the 21-cm trispectrum can be used to include non-Gaussian covariance in an analysis of the 21-cm power spectrum such as [36] (see [37] for a similar analysis in the context of LSS). In principle, non-Gaussian covariance will also impact the forecasts presented in this chapter. We leave this to future work.

- By modeling the effects of the first stars' Lyman-α radiation on the 21-cm brightness temperature, one could extend our analysis from the Dark Ages into the epoch of Cosmic Dawn. Earth-based experiments measuring the 21-cm signal from this epoch are already being prepared and could provide constraints on primordial non-Gaussianity in the near future [38, 39].

Finally, it's imperative to once more mention the practical and technological challenges involved in measuring such a signal. Although the 21-cm signal from the epoch of reionization can be observed from Earth, once we pass into the Dark Ages

2

($z \geq 30$), the signal has redshifted into wavelengths that are largely blocked by Earth's ionosphere, requiring us to design and build a lunar or space-based observatory. Nevertheless, serious efforts are being made to establish a lunar-based observatory with future NASA and ESA moon missions [13, 40, 41]. Furthermore, a host of foreground sources obscure the pristine 21-cm signal, of which synchrotron emission in our galaxy is the most dominant, and the amplitude of these foregrounds increases towards lower frequencies (higher redshifts). In theory, foregrounds can be avoided when looking at fluctuations as the foreground mode appears initially as a line of sight mode on large scales. Interferometers then introduce additional mode coupling that leads to the famous wedge [42–45]. The future of 21-cm as a competitive cosmological probe will rely critically on whether we are capable of extracting the signal away from the wedge (although it might be possible to reconstruct some of the modes in the wedge, see e.g. [46]). However, these challenges are part of the ongoing efforts in this field, both at low redshifts [47–50], reionization redshifts [38, 39, 51–53] and high redshift probes [38, 54, 55].[14] Finally, we have ignored any form of noise due to the experimental setup (e.g. thermal noise, systematics, etc.) as well as any technical challenges such as integration time. We leave a more realistic forecast including these considerations for future work. The main goal of this chapter has been to show the potential of the 21-cm signal at high redshift and to establish the importance of this probe as a way to explore the nature of the early universe in the future. Reaching sensitivities that allow us to answer questions related to new particles and new fields will be limited by the number of available modes [15] and the epoch of the dark ages in principle contains the largest number of these pristine modes.

---

[14]Besides challenges related to signal extraction for the very large interferometers there are also computational challenges which will have to be addressed due to the sheer number of baselines that can be correlated, see e.g. [50] and references therein.

2.A   Forecast Figures

This Appendix collects the results of the forecasts performed for this Chapter.

Total information massive exchange bispectrum
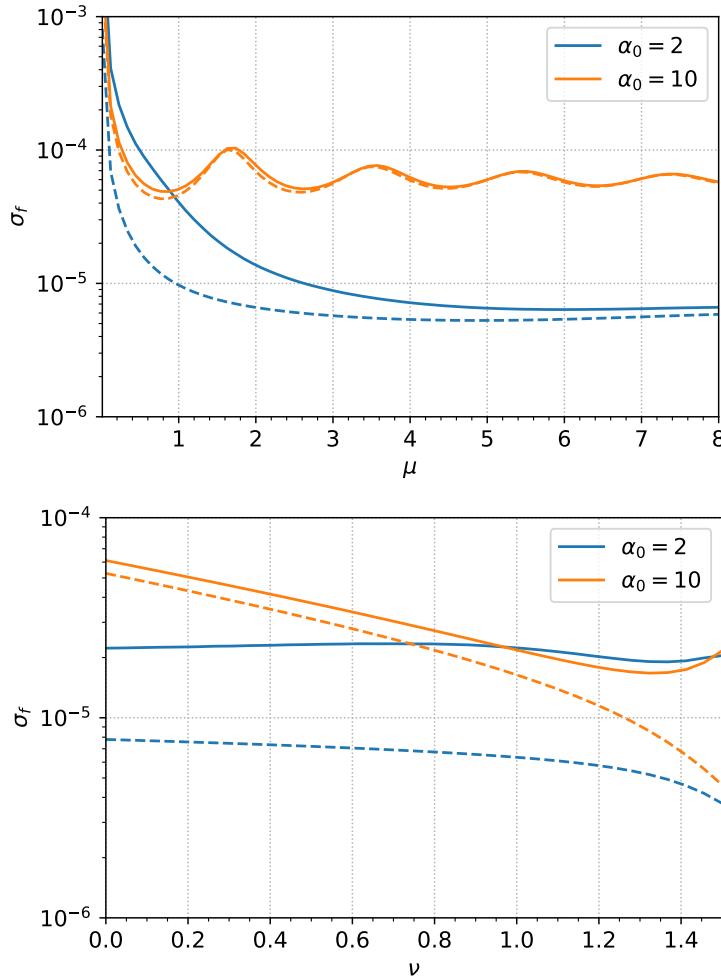
Figure 2.7: *Sensitivity to the amplitude of the massive exchange templates in a cosmic variance limited experiment between $30 \leq z \leq 100$ and $0.01 \leq k \leq 300$ Mpc^{-1}, before (dashed) and after (solid) marginalizing over the common primordial bispectra, for different cutoff values α_0. Top: Clock template, equation (2.56). Bottom: Intermediate template, equation (2.57)*

Template overlap massive exchange bispectrum
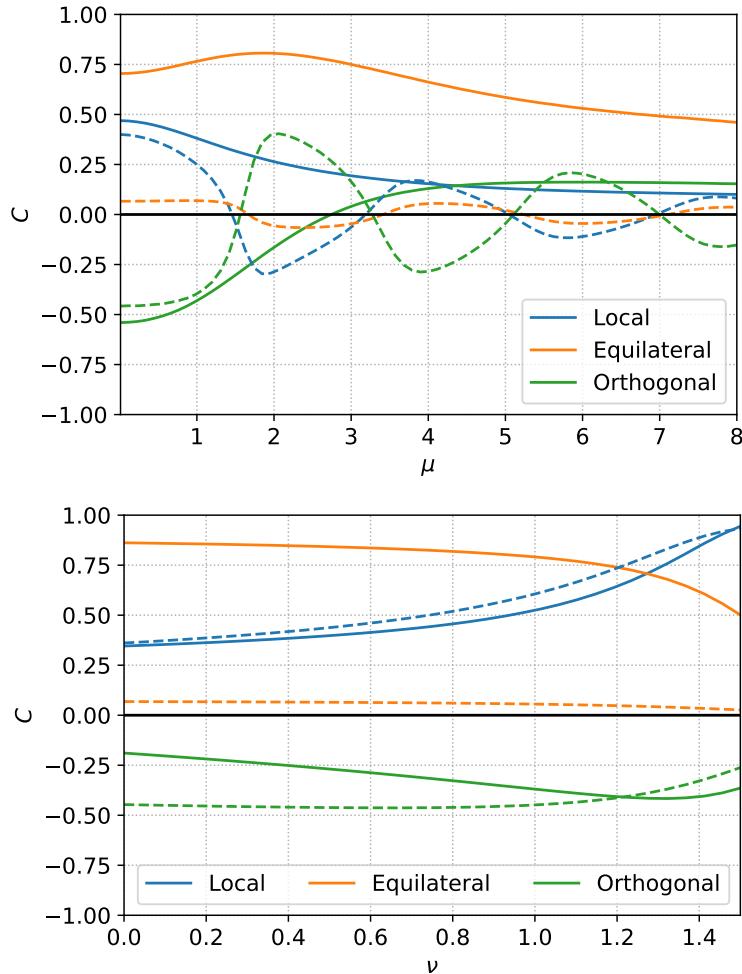
Figure 2.8: *Cosine (2.111) between the massive exchange templates and common primordial shapes for different mass in a cosmic variance limited experiment between* $30 \leq z \leq 100$ *and* $0.01 \leq k \leq 300 \, Mpc^{-1}$. *Solid lines correspond to the template with* $\alpha_0 = 2$. *Dashed lines correspond to the template with* $\alpha_0 = 10$. Top: *Clock template equation (2.56).* Bottom: *Intermediate template equation (2.57)*

Total information massive exchange trispectrum
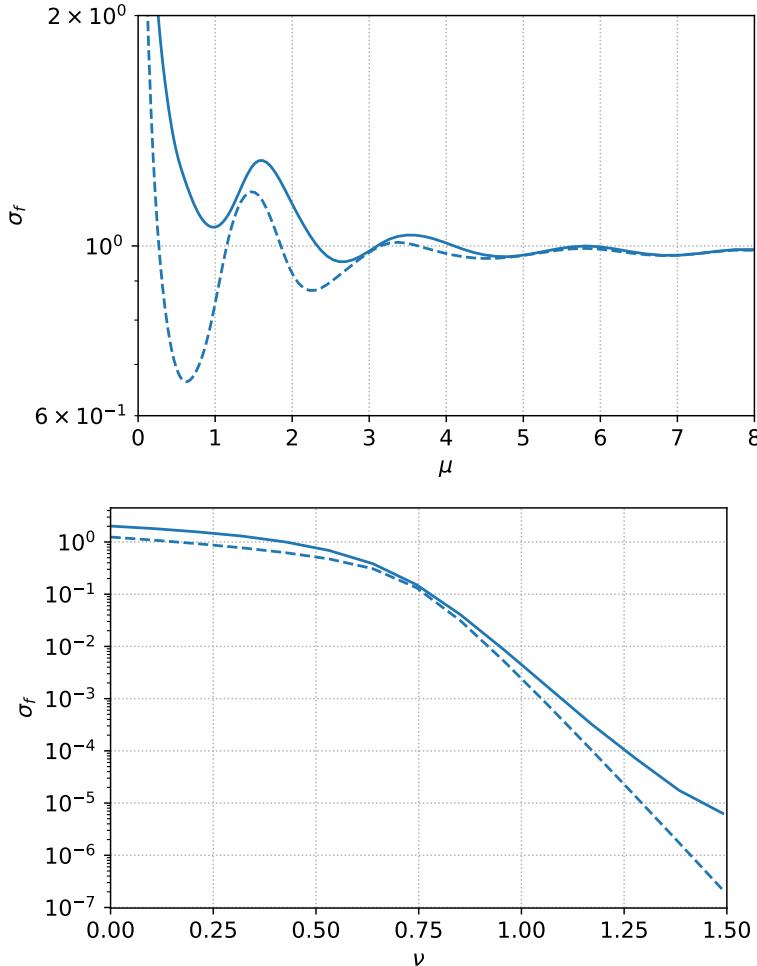
Figure 2.9: *Sensitivity to the amplitude of the exchange trispectrum templates in a cosmic variance limited experiment between* $30 \leq z \leq 100$ *and* $0.01 \leq k \leq 300 \, Mpc^{-1}$, *before (dashed) and after (solid) marginalizing over the common primordial trispectra.* Top: *Clock template equation (2.66).* Bottom: *Intermediate template equation (2.68), here the black line indicates the mass value* $\nu = 3/4$ *at which the enhanced scaling sets in [15].*

2

Template overlap massive exchange trispectrum
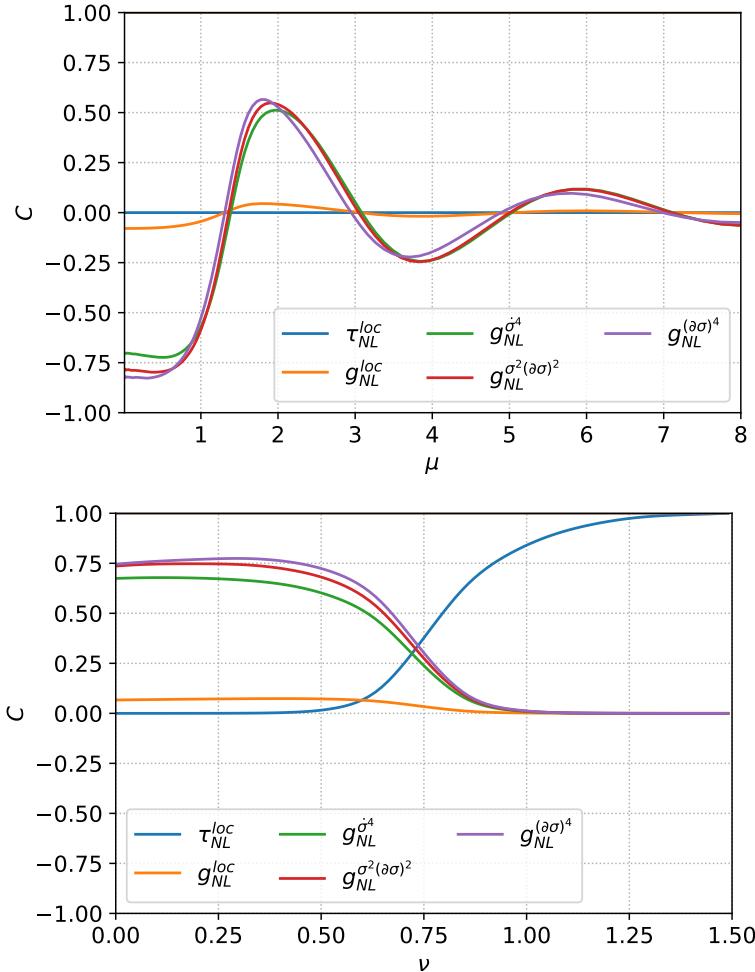
Figure 2.10: *Cosine (2.111) between the exchange templates and the other primordial shapes for different mass in a cosmic variance limited experiment between* $30 \leq z \leq 100$ *and* $0.01 \leq k \leq 300\,Mpc^{-1}$. *Top: Clock template equation (2.66). Bottom: Intermediate template equation (2.68)*

Signal degradation due to secondary bispectrum
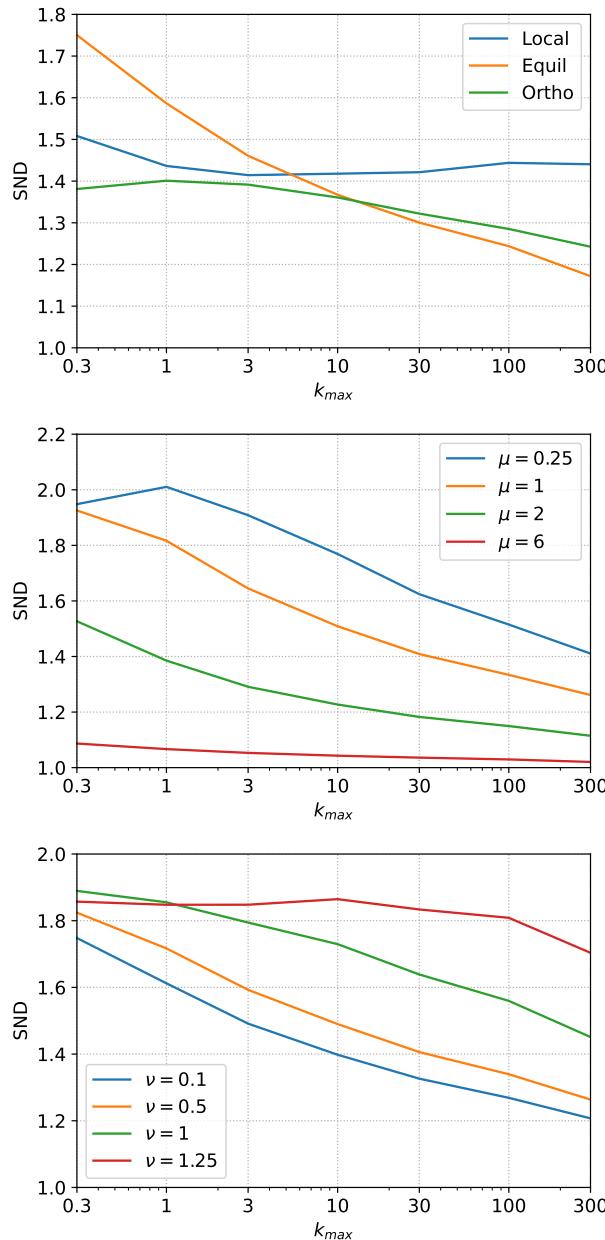
Figure 2.11: *The signal-to-noise degradation factor after marginalizing over the 4 parameter secondary bispectrum as a function of k_{max} for different primordial bispectra. Note the different scales of the vertical axes.*

Sensitivity to clock bispectra as function of $k_{\max}$
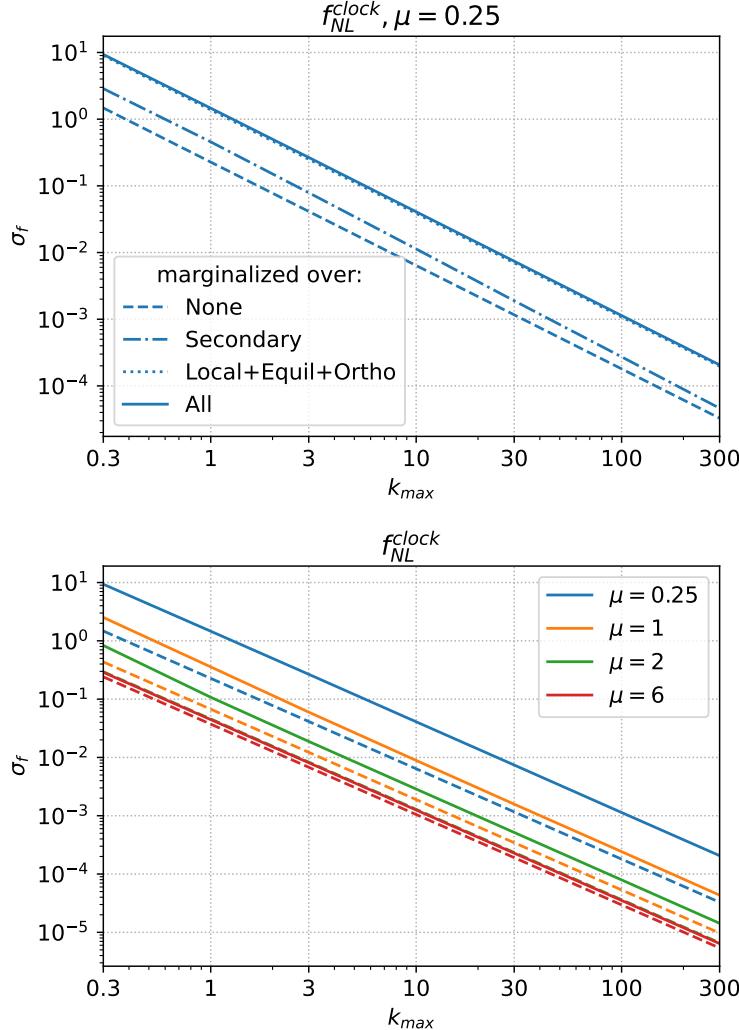
Figure 2.12: Top*: sensitivity to the amplitude of the bispectrum clock template with $\mu = 0.25$ as a function of $k_{\max}$ without marginalizing (dashed), after marginalizing over the 4 parameter secondary bispectrum (dashed-dotted), after marginalizing over the common primordial bispectra (dotted) and after marginalizing over both secondary and common primordial bispectra (solid). Bottom: sensitivity to the amplitude of the clock template for different μ as function of $k_{\max}$ after marginalizing over both secondary and common primordial bispectra.*

Sensitivity to intermediate bispectra as function of k_max
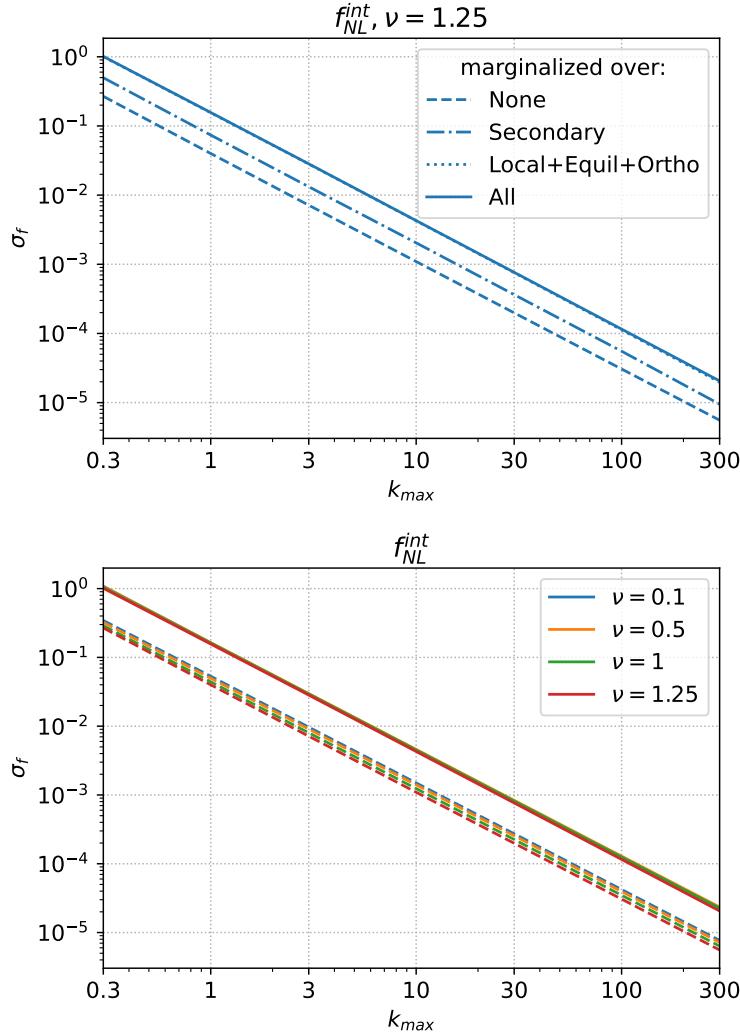
Figure 2.13: Top*: sensitivity to the amplitude of the bispectrum intermediate template with $\nu = 1.25$ as a function of k_max without marginalizing (dashed), after marginalizing over the 4 parameters secondary bispectrum (dashed-dotted), after marginalizing over the common primordial bispectra (dotted) and after marginalizing over both secondary and common primordial bispectra (solid). Bottom: sensitivity to the amplitude of the int template for different ν as function of k_max after marginalizing over secondary and common primordial bispectra.*

Signal degradation due to secondary trispectrum
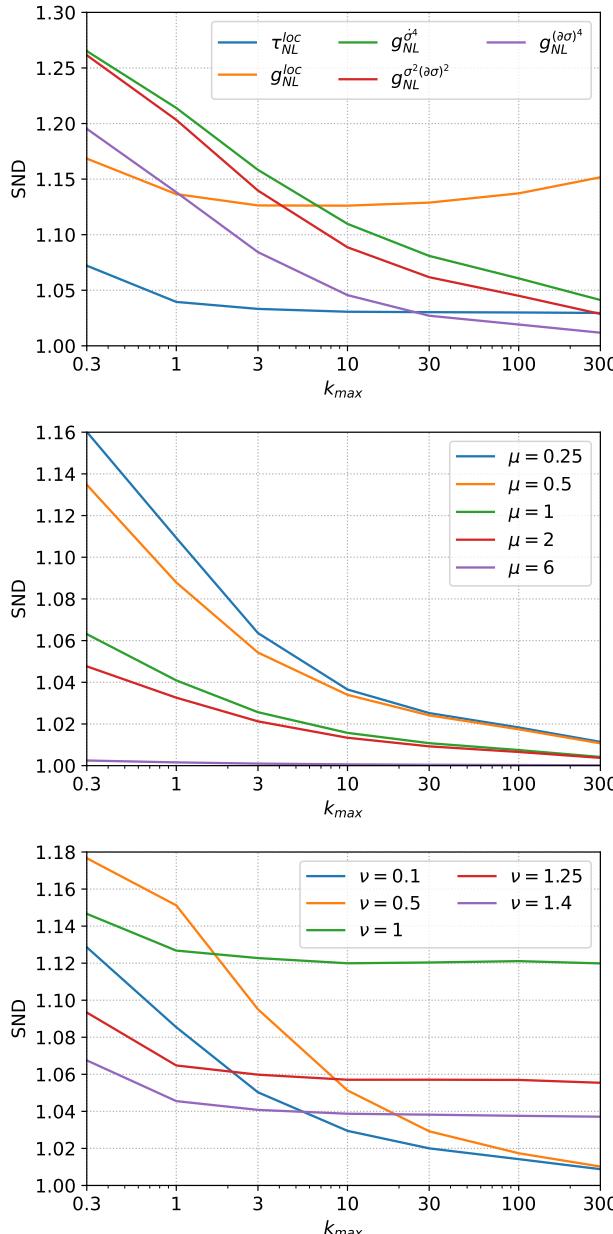
Figure 2.14: *The signal-to-noise degradation factor after marginalizing over the 7 parameter secondary trispectrum as a function of k_{max} for different primordial trispectra. Note the different scales of the vertical axes.*

Sensitivity to clock trispectra as function of k_{max}
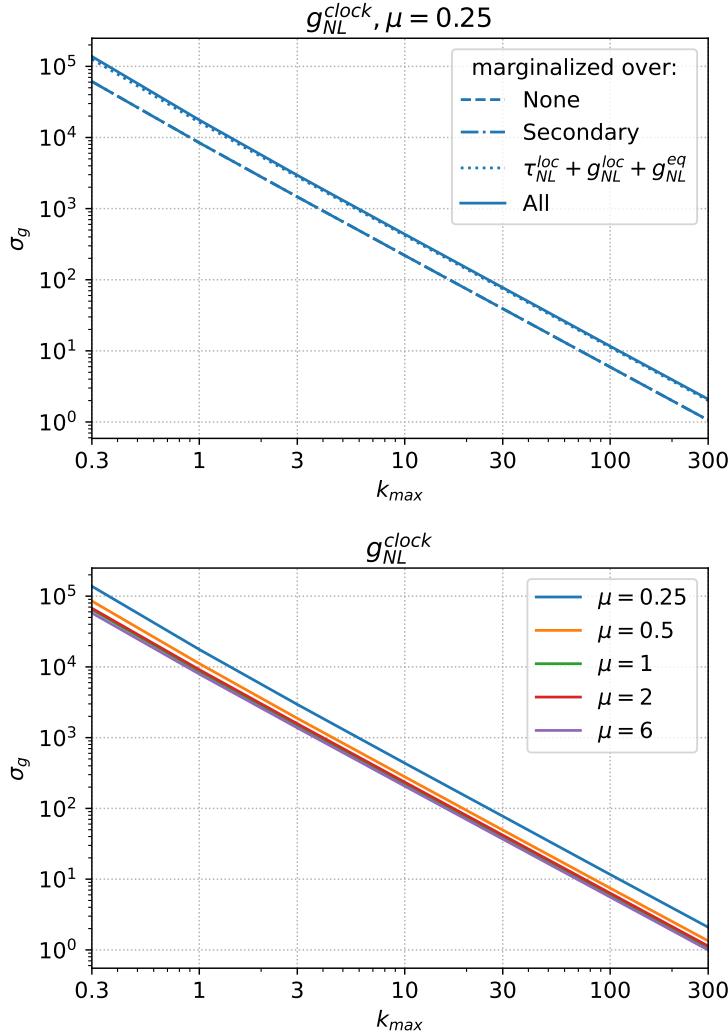
**Figure 2.15:** Top*: sensitivity to the amplitude of the trispectrum clock template with $\mu = 0.25$ as a function of k_{max} without marginalizing (dashed), after marginalizing over the 7 parameter secondary trispectrum (dashed-dotted), after marginalizing over the common primordial trispectra (dotted) and after marginalizing over both secondary and common primordial trispectra (solid). Bottom: sensitivity to the amplitude of the clock template for different μ as function of k_{max} after marginalizing over both secondary and common primordial trispectra.*

Sensitivity to intermediate trispectra as function of k_{max}

Figure 2.16: Top*: sensitivity to the amplitude of the trispectrum intermediate template with $\nu = 0.1$ as a function of k_{max} without marginalizing (dashed), after marginalizing over the 7 parameter secondary trispectrum (dashed-dotted), after marginalizing over the common primordial trispectra (dotted) and after marginalizing over both secondary and common primordial trispectra (solid).* Bottom*: sensitivity to the amplitude of the intermediate template for different ν as function of k_{max} after marginalizing over both secondary and common primordial trispectra.*

Sensitivity to common bispectra as function of baseline/window
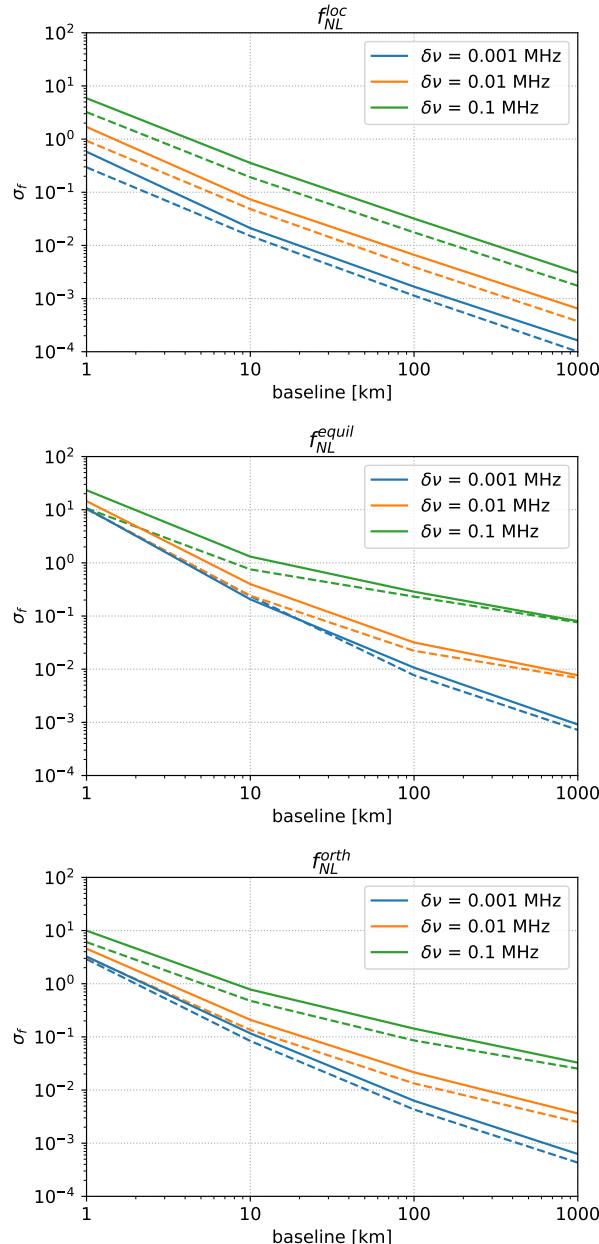

Figure 2.17: *Forecast sensitivity to the amplitude of the common primordial bispectra for different baseline and window size, before (dashed) and after (solid) marginalizing over the 4 parameter secondary bispectrum.*

Sensitivity to clock bispectra as function of baseline/window
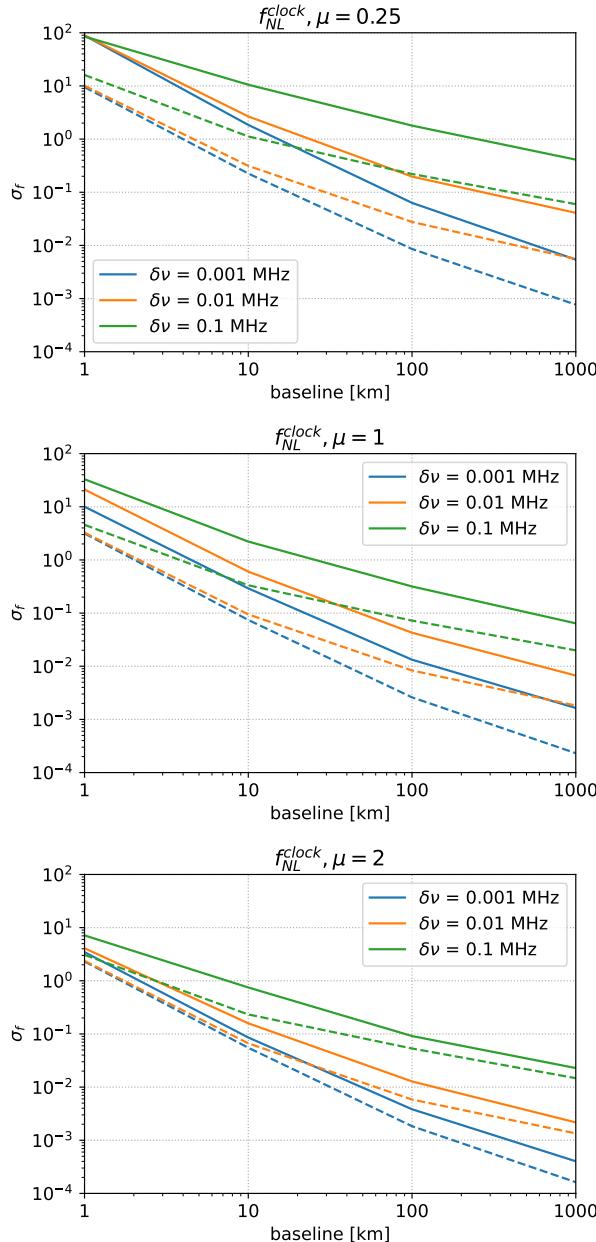
Figure 2.18: *Forecast sensitivity to the amplitude of some massive clock exchange bispectra for different baseline and window, before (dashed) and after (solid) marginalizing over the 4 parameter secondary bispectrum and common primordial bispectra.*

Sensitivity to intermediate bispectra as function of baseline/window
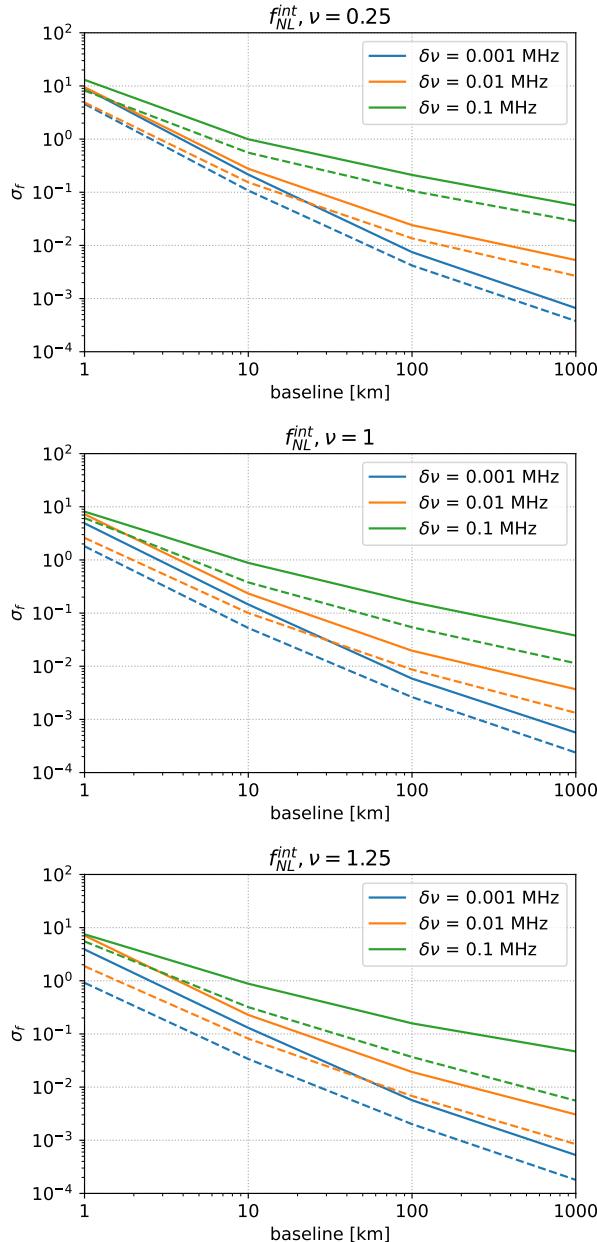
Figure 2.19: *Forecast sensitivity to the amplitude of some massive intermediate exchange bispectra for different baseline and window, before (dashed) and after (solid) marginalizing over the 4 parameter secondary bispectrum and common primordial bispectra.*

Sensitivity to equilateral trispectra as function of baseline/window
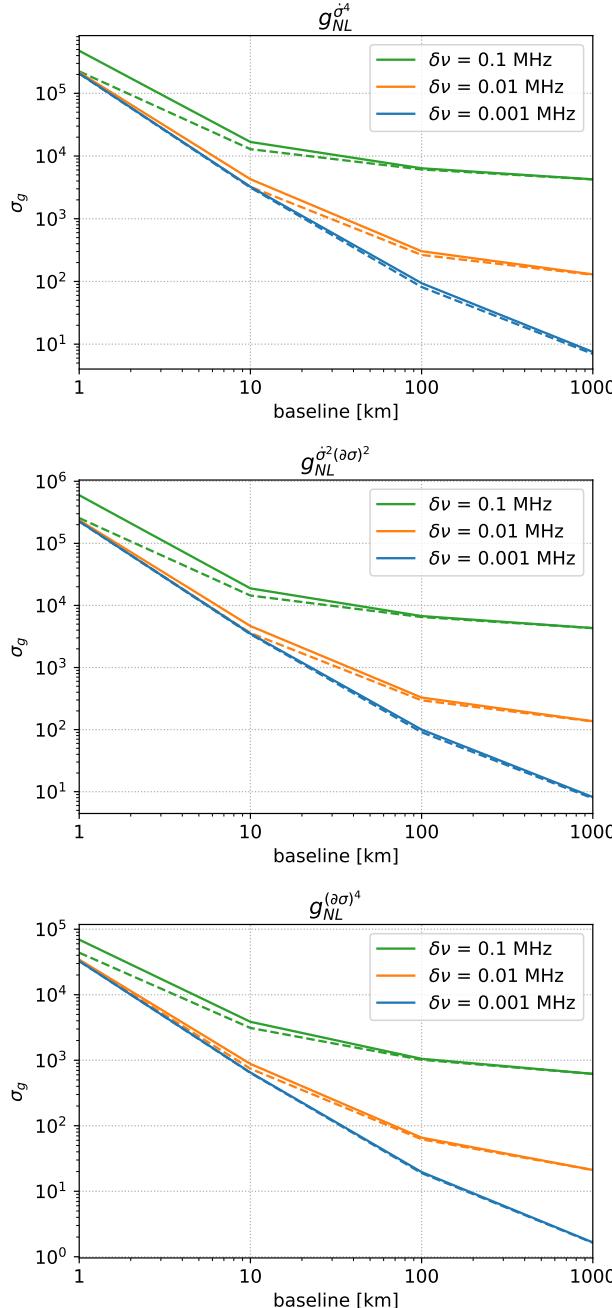
Figure 2.20: *Forecast sensitivity to the amplitude of the primordial trispectra for different baseline and window size, before (dashed) and after (solid) marginalizing over the 7 parameter secondary trispectrum.*

Sensitivity to local trispectra as function of baseline/window
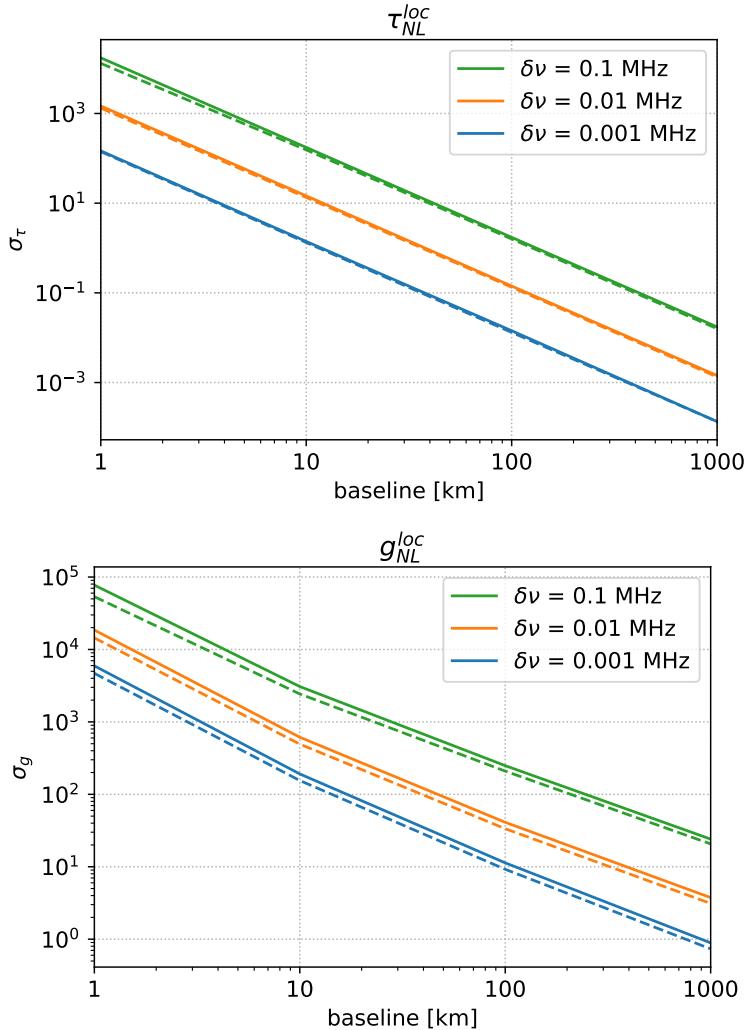
Figure 2.21: *Forecast sensitivity to the amplitude of the primordial trispectra for different baseline and window size, before (dashed) and after (solid) marginalizing over the 7 parameter secondary trispectrum.*

Sensitivity to clock trispectra as function of baseline/window
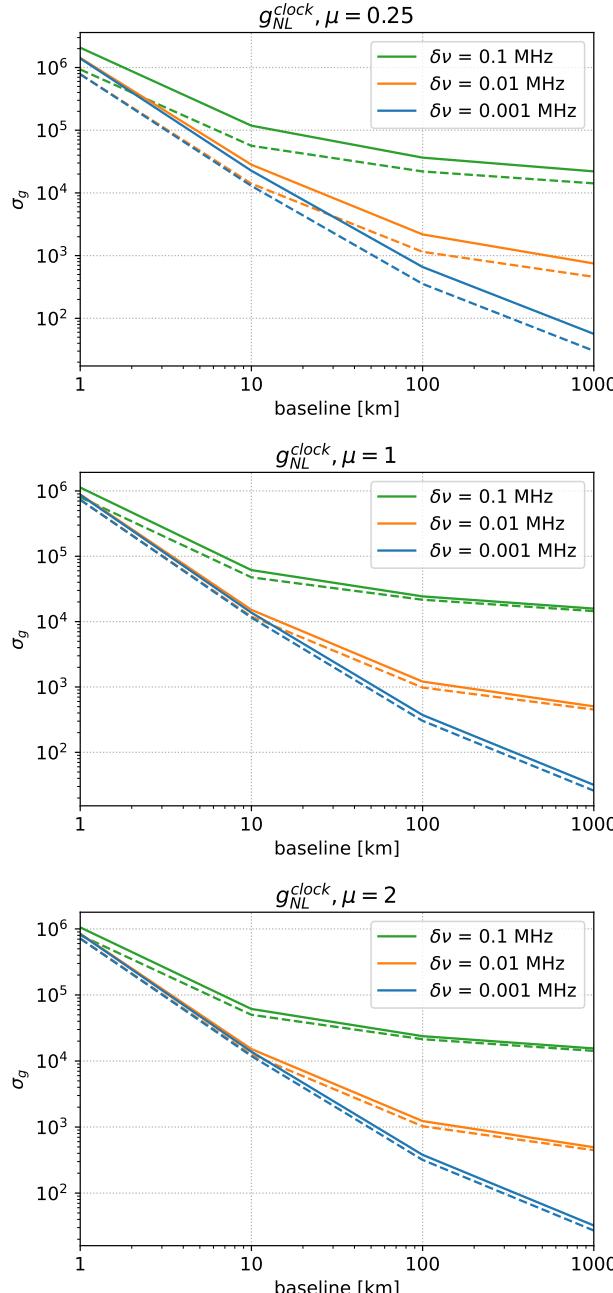
Figure 2.22: *Forecast sensitivity to the amplitude of some clock exchange trispectra for different baseline and window size, before (dashed) and after (solid) marginalizing over the 7 parameter secondary trispectrum and local and equilateral trispectra.*

Sensitivity to intermediate trispectra as function of baseline/window
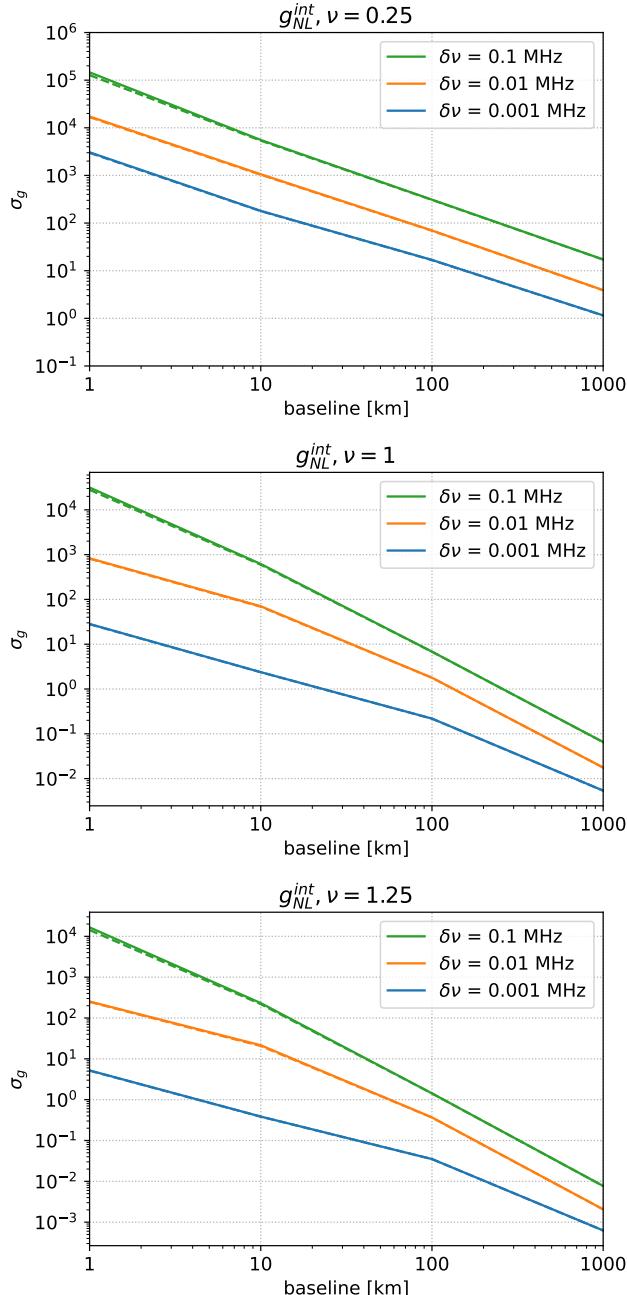
Figure 2.23: *Forecast sensitivity to the amplitude of intermediate massive exchange trispectra for different baseline and window size, before (dashed) and after (solid) marginalizing over the 7 parameter secondary trispectrum and local and equilateral trispectra.*

2.B Evolution of perturbations

In this appendix, we will provide the explicit evolution equations for perturbations in the free electron fraction and gas temperature, used to numerically solve for the coupling coefficients $\mathcal{C}_{i,j}^{x,\mathrm{T}}$ (defined in equation (2.29) and equation (2.30)).

Free electron fraction

To obtain the evolution equations for δ_n^x, we substitute the expansions for δ_{x_e} and δ_{T} (equation (2.28)) into the evolution equation for δ_{x_e} (equation (2.26)). Order by order, we then obtain the following evolution equations:

$$\dot{\delta}_1^x = -\Gamma_{\mathrm{R}}(\delta_1^x + A_1\delta_1^{\mathrm{T}} + \delta_1),$$

$$\dot{\delta}_2^x = -\Gamma_{\mathrm{R}}\Big[\delta_2^x + A_2\delta_2^{\mathrm{T}} + \delta_2 + (\delta_1^x)^2 + 2\delta_1^x\delta_1 + A_1\delta_1^{\mathrm{T}}(\delta_1 + \delta_1^x) + A_2(\delta_1^{\mathrm{T}})^2\Big],$$

$$\dot{\delta}_3^x = -\Gamma_{\mathrm{R}}\Big[\delta_3^x + A_3\delta_3^{\mathrm{T}} + \delta_3 + 2\delta_1^x\delta_2^x + 2(\delta_2^x\delta_1 + \delta_1^x\delta_2) + (\delta_1^x)^2\delta_1$$
$$+ 2A_1(\delta_1^x\delta_2^{\mathrm{T}} + \delta_2^x\delta_1^{\mathrm{T}}) + A_1\delta_1^{\mathrm{T}}\delta_1^x(\delta_1^x + 2\delta_1) + A_1(\delta_1^{\mathrm{T}}\delta_2 + \delta_2^{\mathrm{T}}\delta_1)$$
$$+ A_2(\delta_1^{\mathrm{T}})^2(\delta_1 + 2\delta_1^x) + 2A_2\delta_1^{\mathrm{T}}\delta_2^{\mathrm{T}} + A_3(\delta_1^{\mathrm{T}})^3\Big]. \tag{2.88}$$

Gas temperature

Similarly, we obtain the evolution equations for δ_n^{T} by substituting equation (2.28) into equation (2.20) to get:

$$\dot{\delta}_1^{\mathrm{T}} = \Theta_1,$$

$$\dot{\delta}_2^{\mathrm{T}} = \Theta_2 + \frac{2}{3}H\delta_1(\delta_1^{\mathrm{T}} - \delta_1) - \Gamma_{\mathrm{C}}\delta_1^x\delta_1^{\mathrm{T}},$$

$$\dot{\delta}_3^{\mathrm{T}} = \Theta_3 + \frac{2}{3}H\delta_1\big[(\delta_1)^2 + \delta_2^{\mathrm{T}} - 3\delta_2\big] + \frac{2}{3}H\big[2\delta_2 - (\delta_1)^2\big]\delta_1^{\mathrm{T}} - \Gamma_{\mathrm{C}}[\delta_2^x\delta_1^{\mathrm{T}} + \delta_1^x\delta_2^{\mathrm{T}}], \tag{2.89}$$

where we have used that $\dot{\delta}_n = nH\delta_n$ and defined Θ_n as:

$$\Theta_n \equiv \Gamma_{\mathrm{C}}\Big[(\bar{T}_\gamma/\bar{T}_{\mathrm{gas}} - 1)\delta_n^x - (\bar{T}_\gamma/\bar{T}_{\mathrm{gas}})\delta_n^{\mathrm{T}}\Big] + \frac{2}{3}Hn\delta_n. \tag{2.90}$$

Coupling coefficients

Note that the evolution equations for δ_n^a are coupled *linear* differential equations, we can turn them into coupled differential equations for the coupling coefficients. Compactly, the evolution equations for the couplings can be written as:

$$\frac{d}{da}\mathcal{C}_{n,m}^a = \Delta_{n,m}^a + \Sigma_{n,m}^a, \tag{2.91}$$

in terms $a \equiv (1+z)^{-1}$ as evolution variable. In the compact notation above, $\Delta_{n,m}^a$ denote all terms contributing to the evolution at the order n, whereas $\Sigma_{n,m}^a$ encode the (combined) contribution of lower order perturbations. Explicitly, $\Delta_{n,m}^a$ are defined as:

$$\Delta_{n,m}^{\mathrm{T}} \equiv -\left[n + \frac{\Gamma_{\mathrm{C}}}{H}\frac{\bar{T}_{\gamma}}{\bar{T}_{\mathrm{gas}}}\right]\frac{\mathcal{C}_{n,m}^{\mathrm{T}}}{a} + \frac{\Gamma_{\mathrm{C}}}{H}\left[\frac{\bar{T}_{\gamma}}{\bar{T}_{\mathrm{gas}}} - 1\right]\frac{\mathcal{C}_{n,m}^{x}}{a} + \frac{2n}{3a}\,\delta_{nm},$$

$$\Delta_{n,m}^{x} \equiv -\left[n + \frac{\Gamma_{\mathrm{R}}}{H}\right]\frac{\mathcal{C}_{n,m}^{x}}{a} - \frac{1}{a}\frac{\Gamma_{\mathrm{R}}}{H}\left[A_1\mathcal{C}_{n,m}^{\mathrm{T}} + \delta_{nm}\right]. \tag{2.92}$$

By construction, the evolution of the coefficients $\mathcal{C}_{n,n}^a$ is only sourced by $\mathcal{C}_{n,n}^a$, so that $\Sigma_{n,n}^a = 0$. The cases $n \neq m$ are non-zero and read:

$$\Sigma_{2,1}^{\mathrm{T}} \equiv -\frac{1}{a}\frac{\Gamma_{\mathrm{C}}}{H}\mathcal{C}_{1,1}^{\mathrm{T}}\mathcal{C}_{1,1}^{x} + \frac{2}{3a}\left[\mathcal{C}_{1,1}^{\mathrm{T}} - 1\right],$$

$$\Sigma_{3,1}^{\mathrm{T}} \equiv -\frac{1}{a}\frac{\Gamma_{\mathrm{C}}}{H}\left[\mathcal{C}_{1,1}^{\mathrm{T}}\mathcal{C}_{2,1}^{x} + \mathcal{C}_{2,1}^{\mathrm{T}}\mathcal{C}_{1,1}^{x}\right] + \frac{2}{3a}\left[1 - \mathcal{C}_{1,1}^{\mathrm{T}} + \mathcal{C}_{2,1}^{\mathrm{T}}\right],$$

$$\Sigma_{3,2}^{\mathrm{T}} \equiv -\frac{1}{a}\frac{\Gamma_{\mathrm{C}}}{H}\left[\mathcal{C}_{1,1}^{\mathrm{T}}\mathcal{C}_{2,2}^{x} + \mathcal{C}_{2,2}^{\mathrm{T}}\mathcal{C}_{1,1}^{x}\right] + \frac{2}{3a}\left[-3 + 2\mathcal{C}_{1,1}^{\mathrm{T}} + \mathcal{C}_{2,2}^{\mathrm{T}}\right],$$

$$\Sigma_{2,1}^{x} \equiv -\frac{1}{a}\frac{\Gamma_{\mathrm{R}}}{H}\left[\mathcal{C}_{1,1}^{x}(\mathcal{C}_{1,1}^{x} + 2) + A_1\mathcal{C}_{1,1}^{\mathrm{T}}(1 + 2\mathcal{C}_{1,1}^{x}) + A_2(\mathcal{C}_{1,1}^{\mathrm{T}})^2\right],$$

$$\Sigma_{3,1}^{x} \equiv -\frac{1}{a}\frac{\Gamma_{\mathrm{R}}}{H}\left[2\mathcal{C}_{2,1}^{x}(1 + \mathcal{C}_{1,1}^{x}) + (\mathcal{C}_{1,1}^{x})^2 + A_1\mathcal{C}_{2,1}^{\mathrm{T}}(1 + 2\mathcal{C}_{1,1}^{x})\right.$$
$$\left. + A_1\mathcal{C}_{1,1}^{\mathrm{T}}([\mathcal{C}_{1,1}^{x}]^2 + 2\mathcal{C}_{2,1}^{x} + 2\mathcal{C}_{1,1}^{x})\right]$$
$$- \frac{1}{a}\frac{\Gamma_{\mathrm{R}}}{H}\left[A_2(\mathcal{C}_{1,1}^{\mathrm{T}})^2(1 + 2\mathcal{C}_{1,1}^{x}) + 2A_2\mathcal{C}_{1,1}^{\mathrm{T}}\mathcal{C}_{2,1}^{\mathrm{T}} + A_3(\mathcal{C}_{1,1}^{\mathrm{T}})^3\right],$$

$$\Sigma_{3,2}^{x} \equiv -\frac{1}{a}\frac{\Gamma_{\mathrm{R}}}{H}\left[2\mathcal{C}_{2,2}^{x}(1 + \mathcal{C}_{1,1}^{x}) + 2\mathcal{C}_{1,1}^{x} + A_1\mathcal{C}_{1,1}^{\mathrm{T}}(1 + 2\mathcal{C}_{2,2}^{x})\right.$$
$$\left. + A_1\mathcal{C}_{2,2}^{\mathrm{T}}(1 + 2\mathcal{C}_{1,1}^{x}) + 2A_2\mathcal{C}_{1,1}^{\mathrm{T}}\mathcal{C}_{2,2}^{\mathrm{T}}\right]. \tag{2.93}$$

Initial conditions

To solve the above system of equations, we set the initial conditions following the reasoning of [19]. At high redshifts ($z_i = 1000$) we have $T_{\mathrm{gas}} = T_{\gamma}$ to high accuracy. Since we neglect fluctuations in the photon temperature, we conclude that at z_i gas temperature fluctuations must vanish as well, $\delta_{\mathrm{T}}(z_i, \mathbf{x}) = 0$, or equivalently $\mathcal{C}_{n,m}^{\mathrm{T}}(z_i) = 0$. Regarding electron fraction fluctuations δ_{x_e} (or equivalently $\mathcal{C}_{n,m}^{x}$), one should in principle start computing the evolution at earlier times to get the appropriate initial conditions at the initial redshift z_i. However, we are interested in the effect of δ_{x_e} at late times (during the Dark Ages) on the 21-cm signal, which only enters via the coupling to gas temperature fluctuations. At late times, this coupling

is insensitive to the initial value $\delta_{x_e}(z_i)$, since perturbations grow after recombination and the initial value is quickly forgotten [19]. Therefore, we may set $\delta_{x_e}(z_i) = 0$, corresponding to $\mathcal{C}^x_{n,m}(z_i) = 0$.

2

2.C Fisher information matrix

We summarize the relevant equations used to perform the forecasts in this chapter. Following [36] consider a cosmic variance limited survey over 14 redshift bins between $30 \leq z \leq 100$, assuming these are sufficiently uncorrelated. For the 21-cm bispectrum, the (cosmic variance limited) Fisher information matrix for a redshift bin z_i is defined by [12, 56, 57]:

$$F_{\alpha\beta}(z_i) = \int_{\mathbf{k}_1 \geq \mathbf{k}_2 \geq \mathbf{k}_3} \frac{\partial B_{\delta T}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, z_i)}{\partial p_\alpha} \frac{(2\pi)^6 \delta_D^2(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3)}{P_{\delta T}(\mathbf{k}_1, z_i) P_{\delta T}(\mathbf{k}_2, z_i) P_{\delta T}(\mathbf{k}_3, z_i)}$$
$$\times \frac{\partial B_{\delta T}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, z_i)}{\partial p_\beta}$$

(2.94)

where p_α, p_β are the bispectrum parameters of interest. The total Fisher information is then simply given by the sum over redshift bins. When considering primordial bispectra one sees that all z and line-of-sight momentum dependence in the numerator cancel against the denominator and we find the simplified Fisher information matrix:

$$F_{\alpha\beta} = V_{\text{tot}} \int_{\mathbf{k}_1 \geq \mathbf{k}_2 \geq \mathbf{k}_3} \frac{\partial B_\zeta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)}{\partial p_\alpha} \frac{(2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3)}{P_\zeta(k_1) P_\zeta(k_2) P_\zeta(k_3)} \frac{\partial B_\zeta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)}{\partial p_\beta}$$

(2.95)

where we used $\delta_D(\mathbf{0}) = V/(2\pi)^3$ where V_{tot} is the total comoving survey volume. Now no binning of redshift is necessary. Furthermore, for (isotropic) primordial bispectra, the fisher matrix can be further simplified by integrating out the delta function, to read [58]

$$F_{\alpha\beta} = \frac{V_{\text{tot}}}{8\pi^4} \int_{k_{\min}}^{k_{\max}} dk_1 \int_{k_1/2}^{k_1} dk_2 \int_{k_{\min}^*}^{k_2} dk_3 \frac{\partial B_\zeta(k_1, k_2, k_3)}{\partial p_\alpha} \frac{k_1 k_2 k_3}{P_\zeta(k_1) P_\zeta(k_2) P_\zeta(k_3)}$$
$$\times \frac{\partial B_\zeta(k_1, k_2, k_3)}{\partial p_\beta}$$

(2.96)

where $k_{\min}^* = \text{Max}(k_{\min}, k_1 - k_2)$. Using this expression we determine the total amount of information in 2.4. Contrary to the primordial bispectrum, the 21-cm secondary bispectrum (see e.g. (2.79)) depends explicitly on the line-of-sight momentum through the angles $\mu(\mathbf{k}) = k_\parallel/k$. Furthermore, the redshift dependence in the numerator and denominator of the Fisher matrix equation no longer cancel. Starting from equation (2.94) and integrating out the delta function we find

$$F_{\alpha\beta}(z_i) = \frac{1}{3!} \frac{V_i}{(2\pi)^6} \int_{k_{\min}}^{k_{\max}} d^2\mathbf{k}_{1\perp} \, d^2\mathbf{k}_{2\perp} \, dk_{1\|} \, dk_{2\|} \frac{\partial B_{\delta T}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, z_i)}{\partial p_\alpha}$$

$$\times \left(\prod_{j=1}^{3} \frac{1}{P_{\delta T}(\mathbf{k}_j, z_i)} \right) \frac{\partial B_{\delta T}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, z_i)}{\partial p_\beta}$$

(2.97)

where now $\mathbf{k}_3 = -\mathbf{k}_1 - \mathbf{k}_2$ and the factor $1/3!$ in front is there to count every triangle configuration only once. Going to cylindrical coordinates, we can get rid of one more (angular) degree of freedom by writing $k_{3\perp}^2 = k_{1\perp}^2 + k_{2\perp}^2 + 2k_{1\perp}k_{2\perp}\cos(\phi_{12})$ and invoking isotropy in the perpendicular direction:

$$\begin{aligned} F_{\alpha\beta}(z_i) &= \frac{V_i}{3!(2\pi)^5} \int_0^{2\pi} d\phi_{12} \int_{k_{\min}}^{k_{\max}} dk_{1\perp} \, dk_{2\perp} \, dk_{1\|} \, dk_{2\|} \, k_{1\perp}k_{2\perp} \\ &\times \frac{\partial B_{\delta T}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, z_i)}{\partial p_\alpha} \left(\prod_{j=1}^{3} \frac{1}{P_{\delta T}(\mathbf{k}_j, z_i)} \right) \frac{\partial B_{\delta T}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, z_i)}{\partial p_\beta} \end{aligned}$$

(2.98)

We use this expression to evaluate the Fisher matrix for the 21-cm secondary bispectrum. The integration ranges can be straightforwardly modified to distinguish between $k_{\max}^{\|}$ and $k_{\max}^{\perp}$ as we do in section 2.4.3.

Analogous to the bispectrum, the Fisher matrix for the trispectrum is defined as:

$$\begin{aligned} F_{\alpha\beta}(z_i) = & \int_{\mathbf{k}_1 \geq \mathbf{k}_2 \geq \mathbf{k}_3 \geq \mathbf{k}_4} \frac{\partial T_{\delta T}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, z_i)}{\partial p_\alpha} \times \\ & \frac{(2\pi)^6 \delta_D^2(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4)}{P_{\delta T}(\mathbf{k}_1, z_i) P_{\delta T}(\mathbf{k}_2, z_i) P_{\delta T}(\mathbf{k}_3, z_i), P_{\delta T}(\mathbf{k}_4, z_i)} \frac{\partial T_{\delta T}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, z_i)}{\partial p_\beta} \end{aligned}$$

(2.99)

When considering only primordial trispectra, direction- and z-dependence again drop out. The quadrilateral can then be parametrized by the length of four sides and two diagonals and the Fisher matrix can be shown to take the following form [58]:

$$\begin{aligned} F_{\alpha\beta} = & \frac{V_{\text{tot}}}{4!(2\pi)^3 2\pi^4} \int_{\mathcal{V}_T} dk_1 \, dk_2 \, dk_3 \, dk_4 \, ds \, dt \, \frac{\partial T_\zeta(k_1, k_2, k_3, k_4, s, t)}{\partial p_\alpha} \times \\ & \frac{k_1 k_2 k_3 k_4 st}{\sqrt{g_1} P_\zeta(k_1) P_\zeta(k_2) P_\zeta(k_3) P_\zeta(k_4)} \frac{\partial T_\zeta(k_1, k_2, k_3, k_4, s, t)}{\partial p_\beta} \end{aligned}$$

(2.100)

where $\mathcal{V}_T$ is the tetrahedral region spanned by the quadrilateral, which can be enforced by triangle conditions on every side of the tetrahedron [59]. Furthermore, the function g_1 is given by [58]:

$$g_1 = s^2 t^2 \left(\sum_i k_i^2 - s^2 - t^2 \right) - s^2 \kappa_{23} \kappa_{14} + t^2 \kappa_{12} \kappa_{32} - (k_1^2 k_3^2 - k_2^2 k_4^2)(\kappa_{12} + \kappa_{34})$$

$$(2.101)$$

. Using equation (2.100) we forecast the amount of information available to constrain the trispectrum from the Dark Ages in section 2.4. For the 21-cm secondary trispectrum, we again distinguish between perpendicular and line-of-sight momenta. One way to parametrize these quadrilaterals is using three of its sides:

$$
\begin{aligned}
F_{\alpha\beta}(z_i) &= \frac{1}{4!} \frac{V_i}{(2\pi)^9} \int d^2\mathbf{k}_{1\perp} d^2\mathbf{k}_{2\perp} d^2\mathbf{k}_{3\perp} dk_{1\parallel} dk_{2\parallel} dk_{3\parallel} \\
&\quad \frac{\partial T_{\delta T}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, z_i)}{\partial p_\alpha} \left(\prod_{j=1}^4 \frac{1}{P_{\delta T}(\mathbf{k}_j, z_i)} \right) \frac{\partial T_{\delta T}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, z_i)}{\partial p_\beta}
\end{aligned}
$$

$$(2.102)$$

where it is now understood that $\mathbf{k}_4 = -\mathbf{k}_1 - \mathbf{k}_2 - \mathbf{k}_3$. Furthermore, in cylindrical coordinates we can write

$$
\begin{aligned}
k_{4\perp}^2 = k_{1\perp}^2 + k_{2\perp}^2 + k_{3\perp}^2 &+ 2k_{1\perp}k_{2\perp}\cos(\phi_1 - \phi_2) + 2k_{1\perp}k_{3\perp}\cos(\phi_1 - \phi_3) \\
&+ 2k_{2\perp}k_{3\perp}\cos(\phi_2 - \phi_3) \quad (2.103)
\end{aligned}
$$

and we can get rid of one angular degree of freedom by isotropy, leading to the following expression for the Fisher matrix:

$$
\begin{aligned}
F_{\alpha\beta}(z_i) &= \frac{V_i}{4!(2\pi)^8} \int_0^{2\pi} d\phi_2 \, d\phi_3 \int_{k_{\min}}^{k_{\max}} dk_{1\perp} \, dk_{2\perp} \, dk_{3\perp} \, dk_{1\parallel} \, dk_{2\parallel} \, dk_{3\parallel} \, k_{1\perp} k_{2\perp} k_{3\perp} \\
&\quad \times \frac{\partial T_{\delta T}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, z_i)}{\partial p_\alpha} \left(\prod_{j=1}^4 \frac{1}{P_{\delta T}(\mathbf{k}_j, z_i)} \right) \frac{\partial T_{\delta T}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, z_i)}{\partial p_\beta}
\end{aligned}
$$

$$(2.104)$$

Using this expression we evaluate the Fisher information matrix including the 21-cm secondary trispectra. Note that the integration ranges can be modified to distinguish between $k_{\max}^{\parallel}$ and $k_{\max}^{\perp}$ as we do in section 2.4.3.

Some of the trispectra considered in this work, such as the $\tau_{\mathrm{NL}}^{\mathrm{loc}}$ and intermediate shape, have a diverging behaviour in the collapsed limit, where one of the diagonals becomes very small. In order for the VEGAS integration algorithm [60] to be able

to properly evaluate the Fisher matrix for these trispectra, it is more favorable to parametrize the quadrilateral in terms of one side and two diagonals instead, leading to the alternative expression:

$$
\begin{aligned}
F_{\alpha\beta}(z_i) &= \frac{V_i}{4!(2\pi)^8} \int_0^{2\pi} d\phi_s \, d\phi_t \int_{k_{\min}}^{k_{\max}} dk_{1\perp} \, ds_\perp \, dt_\perp \, dk_{1\parallel} \, ds_\parallel \, dt_\parallel \, k_{1\perp} s_\perp t_\perp \\
&\times \frac{\partial T_{\delta T}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, z_i)}{\partial p_\alpha} \left(\prod_{j=1}^4 \frac{1}{P_{\delta T}(\mathbf{k}_j, z_i)} \right) \frac{\partial T_{\delta T}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, z_i)}{\partial p_\beta}
\end{aligned}
$$

$$(2.105)$$

where the sides of the quadrilateral now read:

$$
\begin{aligned}
k_{2\perp}^2 &= |\mathbf{s}_\perp - \mathbf{k}_{1\perp}|^2 = s_\perp^2 + k_{1\perp}^2 - 2s_\perp k_{1\perp} \cos(\phi_1 - \phi_s), \\
k_{2\parallel} &= s_\parallel - k_{1\parallel} \\
k_{3\perp}^2 &= |\mathbf{k}_{1\perp} - \mathbf{s}_\perp - \mathbf{t}_\perp|^2 = k_{1\perp}^2 + s_\perp^2 + t_\perp^2 - 2k_{1\perp} s_\perp \cos(\phi_1 - \phi_s) \\
&\quad - 2k_{1\perp} s_\perp \cos(\phi_1 - \phi_s) - 2s_\perp t_\perp \cos(\phi_s - \phi_t) \\
k_{3\parallel} &= k_{1\parallel} - s_\parallel - t_\parallel \\
k_{4\perp}^2 &= |\mathbf{t} - \mathbf{k}_1|^2 = t_\perp^2 + k_{1\perp}^2 - 2t_\perp k_{1\perp} \cos(\phi_1 - \phi_t), \\
k_{2\parallel} &= s_\parallel - k_{1\parallel}
\end{aligned}
$$

$$(2.106)$$

Once the Fisher information matrix has been evaluated, the error with which we can determine the parameter p_α is then given by:

$$
\sigma_{p_\alpha} = F_{\alpha\alpha}^{-1/2} \tag{2.107}
$$

while the signal-to-noise is given by its inverse:

$$
\left(\frac{S}{N} \right)_{p_\alpha} = \sqrt{F_{\alpha\alpha}} \tag{2.108}
$$

When trying to constrain multiple parameters the error for each parameters is instead determined by *marginalizing* over the parameter space, which in practise means inverting the Fisher information matrix to obtain the marginalised error:

$$
\sigma_{p_\alpha}^{\mathrm{marg}} = \sqrt{(F^{-1})_{\alpha\alpha}}. \tag{2.109}
$$

The amount of signal that is lost due to the marginalisation can then be quantified by the signal-to-noise degradation (SND) factor:[15]

$$
\mathrm{SND}_{\mathbf{p}_\alpha} = \frac{\sigma_{p_\alpha}^{\mathrm{marg}}}{\sigma_{p_\alpha}} = \sqrt{(F^{-1})_{\alpha\alpha} F_{\alpha\alpha}}. \tag{2.110}
$$

---

[15]Note that we use a slightly different definition of the SND factor as compared to [12], where the SND factor is defined as $\mathrm{SND} = \sqrt{(F^{-1})_{\alpha\alpha} F_{\alpha\alpha} - 1}$.

which depends on how much parameters are correlated. This correlation can be quantified by the Fisher information matrix elements for the parameters considered:

$$C_{\alpha\beta} = \frac{F_{\alpha\beta}}{\sqrt{F_{\alpha\alpha}F_{\beta\beta}}}.$$

(2.111)

which will take a value between -1 and 1 and is often refered to as the *overlap* or *cosine* between two parameters. As an example, for two parameters the SND factor and correlation matrix are then simply related by:

$$\text{SND}_{\mathbf{p}_\alpha} = \sqrt{\frac{1}{1 - (C_{\alpha\beta})^2}}.$$

(2.112)

Hence, an overlap of about $C_{\alpha\beta} = 0.995$ is needed to reduce the sensitivity by a factor of 10.

2

2.D 21-cm secondary trispectrum

We present the explicit expressions of the 11 contributions to the secondary trispectrum of 21-cm brightness temperature fluctuations. The contributions from two second order fluctuations are:

$$
\begin{aligned}
T_{\delta_1\delta_1\delta_2\delta_2} = {}& 4c_1^{(1)}(\mathbf{k}_1)c_1^{(1)}(\mathbf{k}_2)c_1^{(2)}(\mathbf{k}_3)c_1^{(2)}(\mathbf{k}_4)\times \\
& \left[F_2^{(s)}(\mathbf{k}_1,\mathbf{k}_{24})F_2^{(s)}(-\mathbf{k}_2,\mathbf{k}_{24})P_1 P_2 P_{24} + \mathbf{k}_1 \leftrightarrow \mathbf{k}_2 \right] + 5\text{ p.}
\end{aligned}
$$

$$
\begin{aligned}
T_{\delta_1\delta_1\theta_2\theta_2} = {}& 4c_1^{(1)}(\mathbf{k}_1)c_1^{(1)}(\mathbf{k}_2)c_2^{(2)}(\mathbf{k}_3)c_2^{(2)}(\mathbf{k}_4)\times \\
& \left[G_2^{(s)}(\mathbf{k}_1,\mathbf{k}_{24})G_2^{(s)}(-\mathbf{k}_2,\mathbf{k}_{24})P_1 P_2 P_{24} + \mathbf{k}_1 \leftrightarrow \mathbf{k}_2 \right] + 5\text{ p.}
\end{aligned}
$$

$$
\begin{aligned}
T_{\delta_1\delta_1\delta_2\theta_2} = {}& 4c_1^{(1)}(\mathbf{k}_1)c_1^{(1)}(\mathbf{k}_2)c_1^{(2)}(\mathbf{k}_3)c_2^{(2)}(\mathbf{k}_4)\times \\
& \left[F_2^{(s)}(\mathbf{k}_1,\mathbf{k}_{24})G_2^{(s)}(-\mathbf{k}_2,\mathbf{k}_{24})P_1 P_2 P_{24} + \mathbf{k}_1 \leftrightarrow \mathbf{k}_2 \right] + 11\text{ p.}
\end{aligned}
$$

$$
\begin{aligned}
T_{\delta_1\delta_1\delta_2[\delta_1]^2} = {}& 2c_1^{(1)}(\mathbf{k}_1)c_1^{(1)}(\mathbf{k}_2)c_1^{(2)}(\mathbf{k}_3)\times \\
& \left\{ \left[c_3^{(2)}(\mathbf{k}_3,\mathbf{k}_{23}) + c_3^{(2)}(\mathbf{k}_3,-\mathbf{k}_2) \right] F_2^{(s)}(\mathbf{k}_1,\mathbf{k}_{23})P_1 P_2 P_{23} + \mathbf{k}_1 \leftrightarrow \mathbf{k}_2 \right\} \\
& + 11\text{ p.}
\end{aligned}
$$

$$
\begin{aligned}
T_{\delta_1\delta_1\theta_2[\delta_1]^2} = {}& 2c_1^{(1)}(\mathbf{k}_1)c_1^{(1)}(\mathbf{k}_2)c_2^{(2)}(\mathbf{k}_3)\times \\
& \left\{ \left[c_3^{(2)}(\mathbf{k}_3,\mathbf{k}_{23}) + c_3^{(2)}(\mathbf{k}_3,-\mathbf{k}_2) \right] G_2^{(s)}(\mathbf{k}_1,\mathbf{k}_{23})P_1 P_2 P_{23} + \mathbf{k}_1 \leftrightarrow \mathbf{k}_2 \right\} \\
& + 11\text{ p.}
\end{aligned}
$$

$$
\begin{aligned}
T_{\delta_1\delta_1[\delta_1]^2[\delta_1]^2} = {}& c_1^{(1)}(\mathbf{k}_1)c_1^{(1)}(\mathbf{k}_2)\times \\
& \left\{ \left[c_3^{(2)}(\mathbf{k}_3,\mathbf{k}_{13}) + c_3^{(2)}(\mathbf{k}_3,-\mathbf{k}_1) \right]\left[c_3^{(2)}(\mathbf{k}_4,\mathbf{k}_{24}) + c_3^{(2)}(\mathbf{k}_4,-\mathbf{k}_2) \right] \right. \\
& \left. \times P_1 P_2 P_{13} + \mathbf{k}_3 \leftrightarrow \mathbf{k}_4 \right\} + 5\text{ p.}
\end{aligned} \tag{2.113}
$$

The contributions from one third order fluctuation are:

$$T_{\delta_1\delta_1\delta_1\delta_3} = 6c_1^{(1)}(\mathbf{k}_1)c_1^{(1)}(\mathbf{k}_2)c_1^{(1)}(\mathbf{k}_3)c_1^{(3)}(\mathbf{k}_4)\ F_3^{(s)}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)P_1P_2P_3 + 3\ \mathrm{p.}$$

$$T_{\delta_1\delta_1\delta_1\theta_3} = 6c_1^{(1)}(\mathbf{k}_1)c_1^{(1)}(\mathbf{k}_2)c_1^{(1)}(\mathbf{k}_3)c_2^{(3)}(\mathbf{k}_4)\ G_3^{(s)}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)P_1P_2P_3 + 3\ \mathrm{p.}$$

$$T_{\delta_1\delta_1\delta_1[\delta_1]^3} = c_1^{(1)}(\mathbf{k}_1)c_1^{(1)}(\mathbf{k}_2)c_1^{(1)}(\mathbf{k}_3)\left\{\left[c_3^{(3)}(\mathbf{k}_4,-\mathbf{k}_2,-\mathbf{k}_3)+\mathbf{k}_2\leftrightarrow\mathbf{k}_3\right]+2\ \mathrm{c.p.}\right\}$$
$$\times\ P_1P_2P_3 + 3\ \mathrm{p.}$$

$$T_{\delta_1\delta_1\delta_1[\delta_1\delta_2]} = 2c_1^{(1)}(\mathbf{k}_1)c_1^{(1)}(\mathbf{k}_2)c_1^{(1)}(\mathbf{k}_3)\left[c_5^{(3)}(\mathbf{k}_4,\mathbf{k}_{14})F_2^{(s)}(\mathbf{k}_2,\mathbf{k}_3)+2\ \mathrm{c.p.}\right]$$
$$\times\ P_1P_2P_3 + 3\ \mathrm{p.}$$

$$T_{\delta_1\delta_1\delta_1[\delta_1\theta_2]} = 2c_1^{(1)}(\mathbf{k}_1)c_1^{(1)}(\mathbf{k}_2)c_1^{(1)}(\mathbf{k}_3)\left[c_4^{(3)}(\mathbf{k}_4,\mathbf{k}_{14})G_2^{(s)}(\mathbf{k}_2,\mathbf{k}_3)+2\ \mathrm{c.p.}\right]$$
$$\times\ P_1P_2P_3 + 3\ \mathrm{p.} \tag{2.114}$$

where we used the condensed notation $P_i = P_b(\mathbf{k}_i)$ for the baryonic power spectra, c.p. denotes cyclic permutation.

Bibliography

[1] Y. Akrami, F. Arroja, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, and et al., "Planck2018 results," *Astronomy & Astrophysics* **641**, p. A10, Sep 2020.

[2] Y. Akrami *et al.*, "Planck 2018 results. IX. Constraints on primordial non-Gaussianity," *Astron. Astrophys.* **641**, p. A9, 2020.

[3] P. D. Meerburg *et al.*, "Primordial Non-Gaussianity," 3 2019.

[4] J. M. Maldacena, "Non-Gaussian features of primordial fluctuations in single field inflationary models," *JHEP* **05**, p. 013, 2003.

[5] P. Creminelli, "On non-Gaussianities in single-field inflation," *JCAP* **10**, p. 003, 2003.

[6] P. Creminelli and M. Zaldarriaga, "Single field consistency relation for the 3-point function," *JCAP* **10**, p. 006, 2004.

[7] N. Arkani-Hamed and J. Maldacena, "Cosmological Collider Physics," 3 2015.

[8] H. Lee, D. Baumann, and G. L. Pimentel, "Non-Gaussianity as a Particle Detector," *JHEP* **12**, p. 040, 2016.

[9] D. Baumann, G. Goon, H. Lee, and G. L. Pimentel, "Partially Massless Fields During Inflation," *JHEP* **04**, p. 140, 2018.

[10] A. Cooray, "21-cm Background Anisotropies Can Discern Primordial Non-Gaussianity," *Phys. Rev. Lett.* **97**, p. 261301, 2006.

[11] A. Pillepich, C. Porciani, and S. Matarrese, "The bispectrum of redshifted 21-cm fluctuations from the dark ages," *Astrophys. J.* **662**, pp. 1–14, 2007.

[12] P. D. Meerburg, M. Münchmeyer, J. B. Muñoz, and X. Chen, "Prospects for Cosmological Collider Physics," *JCAP* **03**, p. 050, 2017.

[13] J. Silk, "The limits of cosmology: role of the Moon," *Phil. Trans. A. Math. Phys. Eng. Sci.* **379**, p. 20190561, 2021.

[14] J. B. Muñoz, Y. Ali-Haïmoud, and M. Kamionkowski, "Primordial non-gaussianity from the bispectrum of 21-cm fluctuations in the dark ages," *Phys. Rev. D* **92**(8), p. 083508, 2015.

[15] A. Kalaja, P. D. Meerburg, G. L. Pimentel, and W. R. Coulton, "Fundamental limits on constraining primordial non-Gaussianity," *JCAP* **04**, p. 050, 2021.

[16] N. Aghanim *et al.*, "Planck 2018 results. VI. Cosmological parameters," *Astron. Astrophys.* **641**, p. A6, 2020. [Erratum: Astron.Astrophys. 652, C4 (2021)].

[17] S. Furlanetto, S. P. Oh, and F. Briggs, "Cosmology at Low Frequencies: The 21 cm Transition and the High-Redshift Universe," *Phys. Rept.* **433**, pp. 181–301, 2006.

[18] A. Lewis and A. Challinor, "The 21cm angular-power spectrum from the dark ages," *Phys. Rev. D* **76**, p. 083005, 2007.

[19] Y. Ali-Haïmoud, P. D. Meerburg, and S. Yuan, "New light on 21 cm intensity fluctuations from the dark ages," *Phys. Rev. D* **89**(8), p. 083506, 2014.

[20] A. Lewis, "Linear effects of perturbed recombination," *Physical Review D - Particles, Fields, Gravitation and Cosmology* **76**(6), pp. 1–5, 2007.

[21] P. J. E. Peebles, "Recombination of the Primeval Plasma," *Astrophys. J.* **153**, p. 1, 1968.

[22] S. Seager, D. D. Sasselov, and D. Scott, "How exactly did the universe become neutral?," *Astrophys. J. Suppl.* **128**, pp. 407–430, 2000.

[23] C. Cheung, P. Creminelli, A. L. Fitzpatrick, J. Kaplan, and L. Senatore, "The Effective Field Theory of Inflation," *JHEP* **03**, p. 014, 2008.

[24] X. Chen and C. Ringeval, "Searching for Standard Clocks in the Primordial Universe," *JCAP* **08**, p. 014, 2012.

[25] X. Chen, M. H. Namjoo, and Y. Wang, "Models of the Primordial Standard Clock," *JCAP* **02**, p. 027, 2015.

[26] X. Chen and M. H. Namjoo, "Standard Clock in Primordial Density Perturbations and Cosmic Microwave Background," *Phys. Lett. B* **739**, pp. 285–292, 2014.

[27] X. Chen, M. H. Namjoo, and Y. Wang, "Quantum Primordial Standard Clocks," *JCAP* **02**, p. 013, 2016.

[28] X. Chen and Y. Wang, "Large non-Gaussianities with Intermediate Shapes from Quasi-Single Field Inflation," *Phys. Rev. D* **81**, p. 063511, 2010.

[29] T. Suyama and M. Yamaguchi, "Non-Gaussianity in the modulated reheating scenario," *Phys. Rev. D* **77**, p. 023505, 2008.

[30] K. M. Smith, L. Senatore, and M. Zaldarriaga, "Optimal analysis of the CMB trispectrum," 2 2015.

[31] L. Senatore and M. Zaldarriaga, "The Effective Field Theory of Multifield Inflation," *JHEP* **04**, p. 024, 2012.

[32] L. Senatore and M. Zaldarriaga, "A Naturally Large Four-Point Function in Single Field Inflation," *JCAP* **01**, p. 003, 2011.

[33] V. Assassi, D. Baumann, and D. Green, "On Soft Limits of Inflationary Correlation Functions," *JCAP* **11**, p. 047, 2012.

[34] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro, "Large scale structure of the universe and cosmological perturbation theory," *Phys. Rept.* **367**, pp. 1–248, 2002.

2

2

[35] M. Shoji and E. Komatsu, "Third-order Perturbation Theory With Non-linear Pressure," *Astrophys. J.* **700**, pp. 705–719, 2009.

[36] X. Chen, P. D. Meerburg, and M. Münchmeyer, "The Future of Primordial Features with 21 cm Tomography," *JCAP* **09**, p. 023, 2016.

[37] D. Bertolini, K. Schutz, M. P. Solon, J. R. Walsh, and K. M. Zurek, "Non-Gaussian Covariance of the Matter Power Spectrum in the Effective Field Theory of Large Scale Structure," *Phys. Rev. D* **93**(12), p. 123505, 2016.

[38] G. Mellema *et al.*, "Reionization and the Cosmic Dawn with the Square Kilometre Array," *Exper. Astron.* **36**, pp. 235–318, 2013.

[39] D. R. DeBoer *et al.*, "Hydrogen Epoch of Reionization Array (HERA)," *Publ. Astron. Soc. Pac.* **129**(974), p. 045001, 2017.

[40] J. O. Burns, "Transformative science from the lunar farside: observations of the dark ages and exoplanetary systems at low radio frequencies," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **379**, p. 20190564, Nov 2020.

[41] L. V. E. Koopmans *et al.*, "Peering into the dark (ages) with low-frequency space interferometers: Using the 21-cm signal of neutral hydrogen from the infant universe to probe fundamental (Astro)physics," *Exper. Astron.* **51**(3), pp. 1641–1676, 2021.

[42] A. R. Parsons, J. C. Pober, J. E. Aguirre, C. L. Carilli, D. C. Jacobs, and D. F. Moore, "A Per-Baseline, Delay-Spectrum Technique for Accessing the 21cm Cosmic Reionization Signature," *Astrophys. J.* **756**, p. 165, 2012.

[43] A. Liu, A. R. Parsons, and C. M. Trott, "Epoch of reionization window. I. Mathematical formalism," *Phys. Rev. D* **90**(2), p. 023018, 2014.

[44] A. Liu, A. R. Parsons, and C. M. Trott, "Epoch of reionization window. II. Statistical methods for foreground wedge reduction," *Phys. Rev. D* **90**(2), p. 023019, 2014.

[45] A. Ghosh, F. Mertens, and L. V. E. Koopmans, "Deconvolving the wedge: maximum-likelihood power spectra via spherical-wave visibility modelling," *Mon. Not. Roy. Astron. Soc.* **474**(4), pp. 4552–4563, 2018.

[46] C. Modi, M. White, A. Slosar, and E. Castorina, "Reconstructing large-scale structure with neutral hydrogen surveys," *JCAP* **11**, p. 023, 2019.

[47] J. R. Shaw, K. Sigurdson, U.-L. Pen, A. Stebbins, and M. Sitwell, "All-sky interferometry with spherical harmonic transit telescopes," *The Astrophysical Journal* **781**, p. 57, Jan 2014.

[48] X. Chen, "Radio detection of dark energy—the Tianlai project," *Scientia Sinica Physica, Mechanica & Astronomica* **41**, p. 1358, Jan. 2011.

[49] L. B. Newburgh, K. Bandura, M. A. Bucher, T.-C. Chang, H. C. Chiang, J. Cliche, R. Davé, M. Dobbs, C. Clarkson, K. M. Ganga, and et al., "Hirax: a probe of dark energy and radio transients," *Ground-based and Airborne Telescopes VI* , Aug 2016.

[50] A. Slosar *et al.*, "Packed Ultra-wideband Mapping Array (PUMA): A Radio Telescope for Cosmology and Transients," *Bull. Am. Astron. Soc.* **51**, p. 53, 2019.

[51] M. P. van Haarlem *et al.*, "LOFAR: The LOw-Frequency ARray," *Astron. Astrophys.* **556**, p. A2, 2013.

[52] G. B. Taylor, S. W. Ellingson, N. E. Kassim, J. Craig, J. Dowell, C. N. Wolfe, J. Hartman, G. Bernardi, T. Clarke, A. Cohen, N. P. Dalal, W. C. Erickson, B. Hicks, L. J. Greenhill, B. Jacoby, W. Lane, J. Lazio, D. Mitchell, R. Navarro, S. M. Ord, Y. Pihlstrom, E. Polisensky, P. S. Ray, L. J. Rickard, F. K. Schinzel, H. Schmitt, E. Sigman, M. Soriano, K. P. Stewart, K. Stovall, S. Tremblay, D. Wang, K. W. Weiler, S. White, and D. L. Wood, "First light for the first station of the long wavelength array," 2012.

[53] S. J. Tingay *et al.*, "The Murchison Widefield Array: the Square Kilometre Array Precursor at low radio frequencies," *Publ. Astron. Soc. Austral.* **30**, p. 7, 2013.

[54] F. G. Mertens, B. Semelin, and L. V. E. Koopmans, "Exploring the Cosmic Dawn with NenuFAR," in *Semaine de l'astrophysique française 2021*, 9 2021.

[55] C. DiLullo, G. B. Taylor, and J. Dowell, "Using the Long Wavelength Array to Search for Cosmic Dawn," *J. Astron. Inst.* **09**(02), p. 2050008, 2020.

[56] R. Scoccimarro, E. Sefusatti, and M. Zaldarriaga, "Probing primordial non-Gaussianity with large - scale structure," *Phys. Rev. D* **69**, p. 103513, 2004.

[57] T. Baldauf, M. Mirbabayi, M. Simonović, and M. Zaldarriaga, "LSS constraints with controlled theoretical uncertainties," 2 2016.

[58] J. R. Fergusson, D. M. Regan, and E. P. S. Shellard, "Rapid Separable Analysis of Higher Order Correlators in Large Scale Structure," *Phys. Rev. D* **86**, p. 063511, 2012.

[59] X. Chen, B. Hu, M. X. Huang, G. Shiu, and Y. Wang, "Large primordial trispectra in general single field inflation," *Journal of Cosmology and Astroparticle Physics* **2009**, 5 2009.

[60] G. P. Lepage, "Adaptive multidimensional integration: VEGAS enhanced," *J. Comput. Phys.* **439**, p. 110386, 2021.

Chapter 3

Primordial non-Gaussianity and non-Gaussian Covariance

based on

Thomas Flöss, Matteo Biagetti & P. Daniel Meerburg

3

Abstract

In the pursuit of primordial non-Gaussianities, we hope to access smaller scales across larger comoving volumes. At low redshift, the search for primordial non-Gaussianities is hindered by gravitational collapse, to which we often associate a scale k_{NL}. Beyond these scales, it will be hard to reconstruct the modes sensitive to the primordial distribution. When forecasting future constraints on the amplitude of primordial non-Gaussianity, f_{NL}, off-diagonal components are usually neglected in the covariance because these are small compared to the diagonal. We show that the induced non-Gaussian off-diagonal components in the covariance degrade forecast constraints on primordial non-Gaussianity, even when all modes are well within what is usually considered the linear regime. As a testing ground, we examine the effects of these off-diagonal components on the constraining power of the matter bispectrum on f_{NL} as a function of $k_{\max}$ and redshift, confirming our results against N-body simulations out to redshift $z = 10$. We then consider these effects on the hydrogen bispectrum as observed from a PUMA-like 21-cm intensity mapping survey at redshifts $2 < z < 6$ and show that not including off-diagonal covariance over-predicts the constraining power on f_{NL} by up to a factor of 5. For future surveys targeting even higher redshifts, such as Cosmic Dawn and the Dark Ages, which are considered ultimate surveys for primordial non-Gaussianity, we predict that non-Gaussian covariance would severely limit prospects to constrain f_{NL} from the bispectrum.

3.1   Introduction

Over the past few decades, inflation has been established as the leading paradigm
for describing the early universe. It proposes a period of rapidly accelerated expan-
sion during the first fraction of a second after the universe came to be [1–3]. At the
classical level such an expansion can explain why the universe looks nearly identical
in every direction (i.e. is homogeneous and isotropic), while at the quantum level, it
gives rise to the tiny density fluctuations that we observe in the cosmic microwave
background radiation (CMB), which eventually grow into the large scale structure
of the universe (LSS). By precisely mapping the anisotropies in the CMB, we have
determined the fluctuations to be very close to Gaussian distributed, which matches
the predictions of even the simplest theories of inflation [4]. However, in order to sift
through the vast landscape of consistent inflationary theories we are required to look
beyond such general predictions. One avenue to discriminate theories of inflation is
through the study of primordial non-Gaussianities (pnGs) (see [5] and references
therein). Signatures of pnG would appear as non-zero higher n-point functions of
the initial conditions, where the 3-point function, the so-called *bispectrum*, is gener-
ally the most sensitive. A measurement of pnGs can tell us a great deal about the
dynamics driving the expansion (see [6] for a recent overview). To give an example,
a detection of a local-type pNG in future experiments would exclude single-field
models of inflation altogether [7, 8], while generic values of $f_{\mathrm{NL}}^{\mathrm{loc}}$ in multifield mod-
els are $\gtrsim O(1)$ or higher. On the other hand, a detection of equilateral-type pnG of
order $f_{\mathrm{NL}}^{\mathrm{eq}} \geq 1$ would imply strong coupling of the inflaton since it is typically gener-
ated by self-interactions of the inflaton field. Furthermore, particles (fields) present
in the primordial universe leave their unique imprint in the distribution of fluctu-
ations through pnGs, effectively making inflation a particle collider at the highest
conceivable energy scale [9]. Hence, a detailed study of primordial non-Gaussianity
is imperative in order to advance our understanding of the universe as a whole.

While the most stringent constraints on pnGs are derived from measurements of
the CMB bispectrum, future CMB experiments will be limited by its two-dimensional
nature and damping of primary fluctuations. In our search for signatures of pnGs,
we are therefore required to look for alternative probes. Surveys of the large-scale
structure of the universe provide us with a huge observable volume all the way into
the cosmological Dark Ages, by mapping the distribution of galaxies and neutral
hydrogen. While the anisotropies in the CMB are pristine (i.e. linearly related to the
primordial fluctuations), the density field has since evolved. Gravity, being intrinsi-
cally non-linear, breaks the linear relation between density fluctuations and primor-
dial initial conditions, giving rise to a number of complications. Firstly, even if the
primordial fluctuations are purely Gaussian, the non-linear gravitational evolution

introduces *secondary* non-Gaussianities (snGs), typically many orders of magnitude stronger than any primordial signal. Thus, accurate modeling of snG is required in order to properly extract information about pnG. Furthermore, snGs introduce non-Gaussian covariance in the measurements, reducing the amount of *unique* information present in the data. Although the impact of non-Gaussian covariance has been appreciated at low redshifts [10–19], its relevance for high redshift surveys has typically been neglected [20–26]. As upcoming surveys aim to close in on important thresholds such as $f_{\rm NL}^{\rm loc} \sim 1$, it is important to assess the assumptions made in forecasting their performance. In this chapter, we show that by not including non-Gaussian covariance in forecasts of the constraining power of the hydrogen bispectrum observed by a PUMA-like 21-cm intensity mapping experiment [23, 27], one can underestimate the uncertainty in the linear regime by up to a factor of ~ 5 and ~ 2 for the local and equilateral type non-Gaussianity, respectively.

Conventions

We employ the conventions presented at the end of Chapter 1 of this thesis. Additionally, in order to compare to simulations of the matter bispectrum, our cosmology equals the fiducial cosmology of the QUIJOTE suite [28], which closely resembles the 2018 Planck constraints [29]. For the analysis of the PUMA survey, we use the 2015 Planck constraints [30] to match previous forecasts.

3.2   Theoretical Framework and Setup

In order to estimate the signal-to-noise for high-redshift survey observables, we need to introduce a few concepts. We are ultimately interested in constraining the early universe through primordial non-Gaussianities, thus we start off by defining correlations of primordial fluctuations, i.e. our signal of interest. Next, we introduce density perturbations, whose correlations at different positions in the sky are the building blocks of what we actually observe in high- (and low-) redshift surveys. Their dynamics driven by gravity determine the noise we need to overcome.

3.2.1   Initial Conditions

Quantum fluctuations during the inflationary epoch cause the expansion to end at slightly different times in different places, giving rise to tiny scalar density fluctuations ζ that source linear perturbations in the matter density field. In this way, linear fluctuations of the density field trace the primordial initial conditions of the universe. Even a small non-Gaussianity in the distribution of primordial fluctuations serves as an important way to discriminate between different models of inflation. Furthermore, it allows one to directly probe the particle content and interactions of the inflationary epoch [9, 31]. Since such non-Gaussianities are constrained to be small by CMB observations [4], in this chapter we consider only the first non-Gaussian statistic, which is the bispectrum. Hence, we require only the first two statistical moments of the primordial density distributions. In Fourier space, these are the power spectrum $P_\zeta(k)$ and bispectrum $B_\zeta(k_1, k_2, k_3)$, defined as

$$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \rangle \;\; = \;\; (2\pi)^3 \delta_D(\mathbf{k}_{12}) P_\zeta(k_1), \tag{3.1}$$

$$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle \;\; = \;\; (2\pi)^3 \delta_D(\mathbf{k}_{123}) B_\zeta(k_1, k_2, k_3). \tag{3.2}$$

Different inflationary mechanisms give rise to distinct sizes and shapes of bispectra. It is customary to classify these bispectra into three main templates, the so-called local, equilateral, and orthogonal templates, whose expressions are given in Appendix 3.A. The local shape typically arises in models of multi-field inflation and peaks in squeezed triangle configurations $k_1 \ll k_2 \sim k_3$, while the equilateral shape peaks for equilateral configurations $k_1 = k_2 = k_3$. Finally, the orthogonal shape, along with the equilateral one, is a natural prediction of the Effective Field Theory (EFT) of (single field) inflation [32] and peaks for both equilateral and flattened configurations $k_1 = k_2 + k_3$.

3.2.2 Matter field and correlators

The primordial initial conditions serve as the seed for the distribution of matter in the universe. We can therefore study the initial conditions of the universe by studying fluctuations of the matter density field, ρ, defined as $\delta(t, \mathbf{x}) = \rho(t, \mathbf{x})/\bar{\rho}(t) - 1$, with $\bar{\rho}$ the mean density in a volume. Similar to the primordial case, we define correlations of $\delta(\mathbf{x})$ in Fourier space as

$$\langle \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} \rangle = (2\pi)^3 \delta_D(\mathbf{k}_{12}) P_\delta(k_1); \tag{3.3}$$

$$\langle \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} \delta_{\mathbf{k}_3} \rangle = (2\pi)^3 \delta_D(\mathbf{k}_{123}) B_\delta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3), \tag{3.4}$$

$$\langle \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} \delta_{\mathbf{k}_3} \delta_{\mathbf{k}_4} \rangle = (2\pi)^3 \delta_D(\mathbf{k}_{1234}) T_\delta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4), \tag{3.5}$$

where we assume all fluctuations to be at equal times. In this chapter, we also need the 4-point correlation function in Fourier space, known as the *trispectrum*, for the computation of the non-Gaussian covariance. Even in the absence of a primordial bispectrum, or higher-order primordial correlators, fluctuations in the matter field grow via gravitational instability and become non-linear, thereby sourcing the matter bispectrum, trispectrum, and higher-order correlations. The dynamical equations for δ describing this process can be solved perturbatively (see e.g. [33] for a review). This allows one to compute correlators analytically up to a mildly non-linear scale k_{NL}. One way to estimate this scale is by computing

$$k_{\mathrm{NL}}(z) = \left[\frac{1}{6\pi^2} \int_0^\infty dk \, P_\delta^L(k, z) \right]^{-1/2}, \tag{3.6}$$

where P_δ^L is the linear matter power spectrum as defined in Eq. (3.23). We use this scale to confine ourselves to the linear regime [1]. The gravitationally induced bi- and trispectrum in this framework are presented in Appendix 3.B.

To complement the perturbative approach, we resort to N-body simulations of the universe at large scales solving the dynamical equations for δ numerically (see [35] for a review). The advantage of N-body simulations is that they allow us to directly measure correlations of δ even at non-linear scales and to test analytic predictions. The drawback is that they are computationally expensive to run. We make use of publicly available QUIJOTE simulations [28] for our estimates of signal-to-noise at low redshift (i.e. up to $z = 3$). For higher redshifts, as the non-linear scale is pushed to very small scales, instead of fully solving dynamical equations we resort

---

[1]Other definitions have been considered in the literature, e.g. [34] studies the non-linear scale for the bispectrum specifically. The precise definition of k_{NL} does not qualitatively change the results of this chapter.

to `Monofonic` [36], which computes particle positions by solving third-order Lagrangian perturbation theory (3LPT) equations. Further details on how simulation data is used can be found in Appendix 3.C.

3.2.3 Fisher information and estimated uncertainty

In this section, we introduce the quantities we use to estimate the uncertainty on the amplitude of primordial non-Gaussianity, f_{NL}, from observations of the bispectrum.

Fisher matrix

A common way to quantify the information content of an observable is through the Fisher matrix. It encodes both the amount of information available from a measurement to constrain a parameter, as well as the correlation between different parameters. Given N measurements of an observable, which for us will be the matter or hydrogen bispectrum, and a set of parameters we want to constrain, $\mathbf{p}$, the Fisher matrix is defined as

$$F_{ab} = \sum_{TT'} \frac{\partial B_T}{\partial p_a} \, (C)^{-1}_{TT'} \, \frac{\partial B_{T'}}{\partial p_b}, \tag{3.7}$$

where T are triangle configurations in which the bispectra are measured, or calculated, $\mathbf{B}$ is the data vector of bispectra, and $C_{TT'}$ is the covariance of $\mathbf{B}$, defined as

$$C_{TT'} = \langle B_T \, B_{T'} \rangle - \langle B_T \rangle \langle B_{T'} \rangle . \tag{3.8}$$

The estimated uncertainty on a parameter p_a is then defined as

$$\sigma_{p_a} = \left(F^{-1} \right)^{1/2}_{a\,a}, \tag{3.9}$$

where F^{-1} indicates the matrix inverse of F.

N-body measurements

When using numerical simulations, we measure the matter bispectrum on a finite size box with periodic boundary conditions, such that in this case $\delta_{\mathbf{k}}$ is a discrete Fourier transform of the density contrast. The bispectrum estimator then is defined as [37]

$$\hat{B}(k_1, k_2, k_3) \equiv \frac{k_F^3}{N_{tr}} \sum_{\mathbf{q} \in k} \delta_K(\mathbf{q}_{123}) \, \delta_{\mathbf{q_1}} \, \delta_{\mathbf{q_2}} \, \delta_{\mathbf{q_3}}, \tag{3.10}$$

being $k_F = 2\pi/L$ the fundamental frequency in a cubic box of side L, N_{tr} gives the number of "fundamental triangles" formed by the vectors $\mathbf{q}_i$ satisfying the condition $\mathbf{q}_{123} = 0$ that belong in the "triangle bin" defined by the triplet of bin centers (k_1, k_2, k_3) and bin width Δk [2].

The advantage of using N-body simulations is that the full covariance can be estimated numerically from a sample of simulations using Eq. (3.8), where now the average $\langle \cdot \rangle$ is over different realizations of the same simulation. It is also straightforward to compute the Gaussian contribution only, i.e. the case where different modes are uncorrelated. This contribution is given by the product of three power spectra [3]

$$C_{TT'}^G \simeq \frac{(2\pi)^3 \, k_F^3}{V_{123}} \, s_{123} \, \hat{P}(k_1)\hat{P}(k_2)\hat{P}(k_3) \, \delta_{TT'} \tag{3.12}$$

where T, T' denote triangle bins, $V_{123} \simeq 8\pi^2 k_1 k_2 k_3 \Delta k^3$ is the volume of the bin, $s_{123} = 1, 2, 6$ for scalene, isosceles and equilateral triangles respectively and $\hat{P}(k_i)$ are power spectrum measurements.

Limit of infinitely thin bins

At high redshifts, the non-linear scale k_{NL} is pushed to smaller scales. At fixed bin width Δk, this implies a wider range of scales explored, and consequently a larger data vector and covariance. In order to keep the calculations within reasonable computational cost, one solution is to widen the range of bins and sample wavenumbers in log space. Alternatively, we choose to go in the limit of infinitely thin bins and promote the sums to integrals, such that the Fisher matrix becomes

$$F_{ab} = \int_{TT'} \frac{\partial B_T}{\partial p_a} C_{TT'}^{-1} \frac{\partial B_T'}{\partial p_b}, \tag{3.13}$$

where now the matter, or hydrogen, bispectrum is estimated using perturbation theory, as explained in the Appendix. Calculating Eq. (3.13) now implies knowledge of the dependence on triangle configurations T, T' of the inverted full covariance matrix, which is typically hard to compute. In Appendix 3.D we outline a strategy that is based on splitting the covariance into Gaussian and non-Gaussian contributions, $C = C_G + C_{nG}$, and expanding the inverse as a Neumann series. We then

---

[2]We also measure the power spectrum, since, as we show below, it enters in the calculation of the covariance. The estimator of the power spectrum is

$$\hat{P}(k) \equiv \frac{k_F^3}{N_k} \sum_{\mathbf{q} \in k} \delta_{\mathbf{q}} \, \delta_{-\mathbf{q}}, \tag{3.11}$$

where N_k gives the number of modes in each k-bin.

[3]The approximate equality indicates the thin shell approximation.

approximate this series such that the Fisher matrix in the limit of thin bins becomes:

$$F_{ab} = \frac{\left(F_{ab}^{\mathrm{G}}\right)^2}{F_{ab}^{\mathrm{G}} + \delta F_{ab}^{\mathrm{nG}}}, \tag{3.14}$$

where F_{ab}^{G} is the Fisher matrix computed using only the Gaussian covariance C_{G} to compute the inverse covariance and $\delta F_{ab}^{\mathrm{nG}}$ is the non-Gaussian correction computed using as inverse the product of matrices $-C_{\mathrm{G}}^{-1} C_{\mathrm{nG}} C_{\mathrm{G}}^{-1}$.

Model for non-Gaussian covariance

The goal of this chapter is to compute how σ varies for f_{NL} whether we are considering only the Gaussian term C_{G} or a more complete modeling of the covariance including non-Gaussian terms. As explained above, when using N-body simulations, the Gaussian and the full covariance are computed numerically. In the case of thin bins, we need to introduce a model of the bispectrum covariance. Inserting Eq. (3.10) into Eq. (3.8), the computation involves the correlator of 6 fields in Fourier space, which can be combined in four different ways: the Gaussian term is the product of three power spectra ('PPP' term), given by Eq. (3.12). Non-Gaussian terms are represented by either the product of two bispectra ('BB' term), the product of a power spectrum and a trispectrum ('PT' term), or finally the connected 6-point function, the so-called *pentaspectrum*. The pentaspectrum is negligible in most practical cases (see [18] for a rough estimate). The key point of this chapter is to account for the 'BB' and 'PT' terms in signal-to-noise estimates at high redshifts. The 'BB' term, again assuming that correlators are slowly varying in the k-shells, can be written as

$$C_{\mathrm{nG}}^{\mathrm{BB}} \simeq B_T B_{T'} (\Sigma_{TT'}^{11} + 8 \text{ perm}), \tag{3.15}$$

where $\Sigma_{TT'}^{ij}$ is a mode-counting factor that again depends on the shape of the triangles. The 'PT' term is similarly written. We calculate these terms for the matter bispectrum predictions using Eqs. (3.23), (3.22) and (3.24). For the hydrogen bispectrum, we use the following model for the covariance

$$C \approx C_{\mathrm{G}} + 2\, C_{\mathrm{nG}}^{\mathrm{BB}}, \tag{3.16}$$

where the 'PT' term is approximated to be equal to the 'BB' term, which is a good approximation for squeezed triangles for which the non-Gaussian terms are largest [18].

3.3 Constraining f_{NL} at high redshifts

The primary goal of this chapter is to show the importance of including non-Gaussian terms in the covariance when estimating the uncertainty to the primordial non-

Gaussian amplitude f_{NL} in high-redshift surveys.

One could be tempted to neglect the non-Gaussian covariance at high redshifts on scales larger than k_{NL} at that redshift. In this linear regime, one might expect modes of different wavelengths to be mostly uncorrelated, such that the covariance is diagonal and Gaussian terms dominate. As we show in what follows, this intuition fails: off-diagonal terms become important well within what is typically considered the linear regime based on Eq. (3.6).

3.3.1 Uncertainty on f_{NL} from the matter bispectrum

As a testing ground, we first consider the matter bispectrum in real space as our observable and compute the estimated uncertainty of the primordial non-Gaussian amplitude f_{NL} for the primordial bispectra of the local, equilateral, and orthogonal type as defined previously. In this setup, f_{NL} is the only parameter. When using finite-sized bins, we evaluate the derivative $\partial\mathbf{B}/\partial f_{\mathrm{NL}}$ averaging over the bins (see Eq. (3.27) of Appendix 3.C), while in the case of infinitely thin bins, the derivative is analytically computed directly from the templates Eqs. (3.17), (3.18) and (3.19).

Figure 3.1 shows the ratio of the estimated uncertainty computed with non-Gaussian over Gaussian covariance as a function of the maximum wavenumber k_{max}. The uncertainty computed using the infinitely thin bin approximation is shown in solid lines, while simulation measurements are shown as diamonds. Solid lines are computed up to the non-linear scale k_{NL} at that redshift. The uncertainty on local type non-Gaussianity is most affected by the introduction of off-diagonal covariance, increasing by a factor of ~ 5 at $k_{\mathrm{max}} \approx k_{\mathrm{NL}}$ for redshifts lower than $z = 10$ and even higher at higher redshifts. This is because the off-diagonal covariance is largest for squeezed triangle configurations where the local type non-Gaussianity has most of its signal [18] [4]. Equilateral non-Gaussianity is less affected, since most of its signal comes from equilateral triangle configurations whose covariance is large only when approaching non-linear scales. Still, the loss is almost a factor of 2 at $z \lesssim 10$. For a discussion and the results of the orthogonal bispectrum, we refer to Appendix 3.C.

It is important to note that these results do not imply that the uncertainty does not improve overall, since we are still able to access more modes as we increase k_{max}. Rather, our analysis shows that off-diagonal non-Gaussian covariance reduces the amount of information gained by probing smaller scales, or in other words, the signal-to-noise saturates. We show a clear representation of this fact in Figure 3.2,

---

[4]This is very similar to how lensing-induced covariance mostly affects measurements of the local bispectrum in the CMB [38]

where we directly plot the uncertainty of f_{NL} including non-Gaussian terms as a function of k_{max} at different redshifts for a fictitious matter field survey.

3.3.2 Uncertainty on f_{NL} from the hydrogen bispectrum

To make contact with actual future observations, we consider a realistic PUMA-like intensity mapping survey setup. PUMA is a proposed 21-cm intensity mapping experiment aimed at measuring the distribution of neutral hydrogen through the 21-cm hyperfine transition between redshift 2 and 6. One of the key science drivers for PUMA is to provide better constraints on primordial non-Gaussianity with respect to the CMB [27] (see also Figure 5 in [6] for a comparison to other future surveys).

As compared to the simplified scenario considered in Figure 3.1, the calculation of the estimated uncertainty in this case involves several complications. First of all, neutral hydrogen is a biased tracer of the matter field. This introduces additional non-linearities and we need to define a set of nuisance (bias) parameters that are fixed through observations (see [39] for a review). Secondly, we need to compute correlators in redshift space, taking into account the survey geometry and foregrounds. Lastly, the presence of primordial non-Gaussianity introduces additional bias parameters. This last effect famously appears already at the power spectrum level for the local template, known as scale-dependent bias ([40–42] and [43] for a recent review). For this reason, forecasts of $\sigma(f_{\mathrm{NL}})$ depend sensitively on many assumptions, and would need to include the tracer power spectrum in order to be realistic. Here we limit ourselves to calculate the uncertainty using the tracer bispectrum only, rather than performing a full forecast, since our goal is to show the loss of constraining power due to the inclusion of non-Gaussian covariance on the bispectrum [5]. For our computations, we follow the setup presented in [23]. 21-cm intensity mapping is complicated by foregrounds, especially on large scales in the line-of-sight direction. Therefore, the largest scale is effectively set by a foreground cut in the line-of-sight direction ($k_{\parallel,\mathrm{min}} = 0.01$ h/Mpc). The analysis is limited to linear scales by choosing $k_{\mathrm{max}}(z)$ to fractions of $k_{\mathrm{NL}}(z)$, namely $0.5\, k_{\mathrm{NL}}(z)$ and $0.75\, k_{\mathrm{NL}}(z)$, where the non-linear scale is given in Eq. (3.6). We also include the foreground wedge, further reducing the number of modes. Besides f_{NL}, the hydrogen bispectrum is a function of 7 parameters: three bias parameters, two shot-noise parameters, the dimensionless linear growth rate f and the velocity dispersion σ_v. We compute the fiducial value of these parameters as a function of redshift following [23, 45] and the expression for the hydrogen bispectrum is found in the Appendix, Eq. (3.35). We also calculate the hydrogen power spectrum, given in Eq. (3.34), as we use it to compute the Gaus-

---

[5]We have confirmed that our forecasts, using very similar assumption about the PUMA survey, result in forecasts on $\sigma(f_{\mathrm{NL}})$ that are consistent with those presented in Refs. [23, 44] when neglecting non-Gaussian covariance
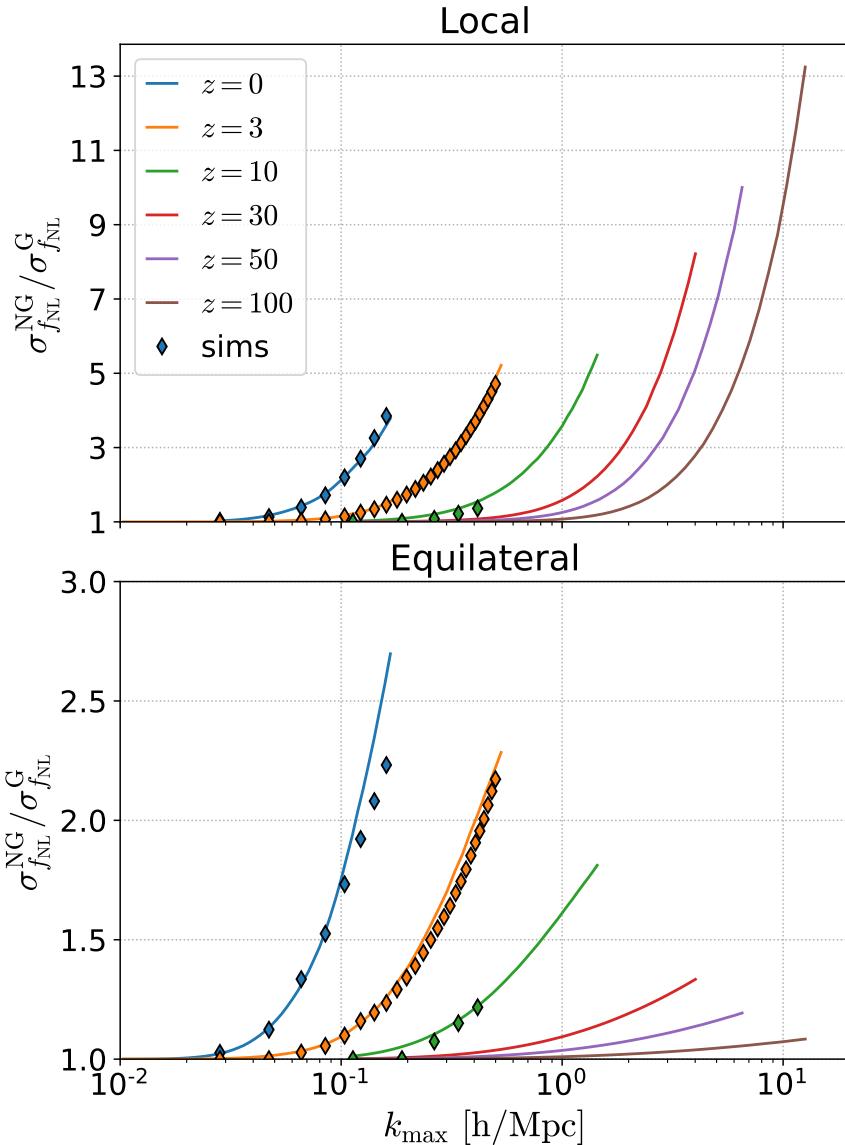
Figure 3.1: *Estimate of relative increase in error on* f_{NL} *due to non-Gaussian covariance as a function of* $k_{\max}$*. The diamonds present results obtained using the* QUIJOTE *simulations* ($z = 0, 3$) *or 3LPT* ($z = 10$)*. Note the different scales on the vertical axes. The local bispectrum is expected to be significantly affected when accounting for non-diagonal covariance even at very high redshifts. Solid lines are estimated up to the non-linear scale* k_{NL} *at each redshift. For* $z = 0$ *and* 3 *the simulation results are also shown up to the non-linear scale, while for* $z = 10$ *they are shown up to the scale at which shot noise becomes a significant contribution to the covariance.*
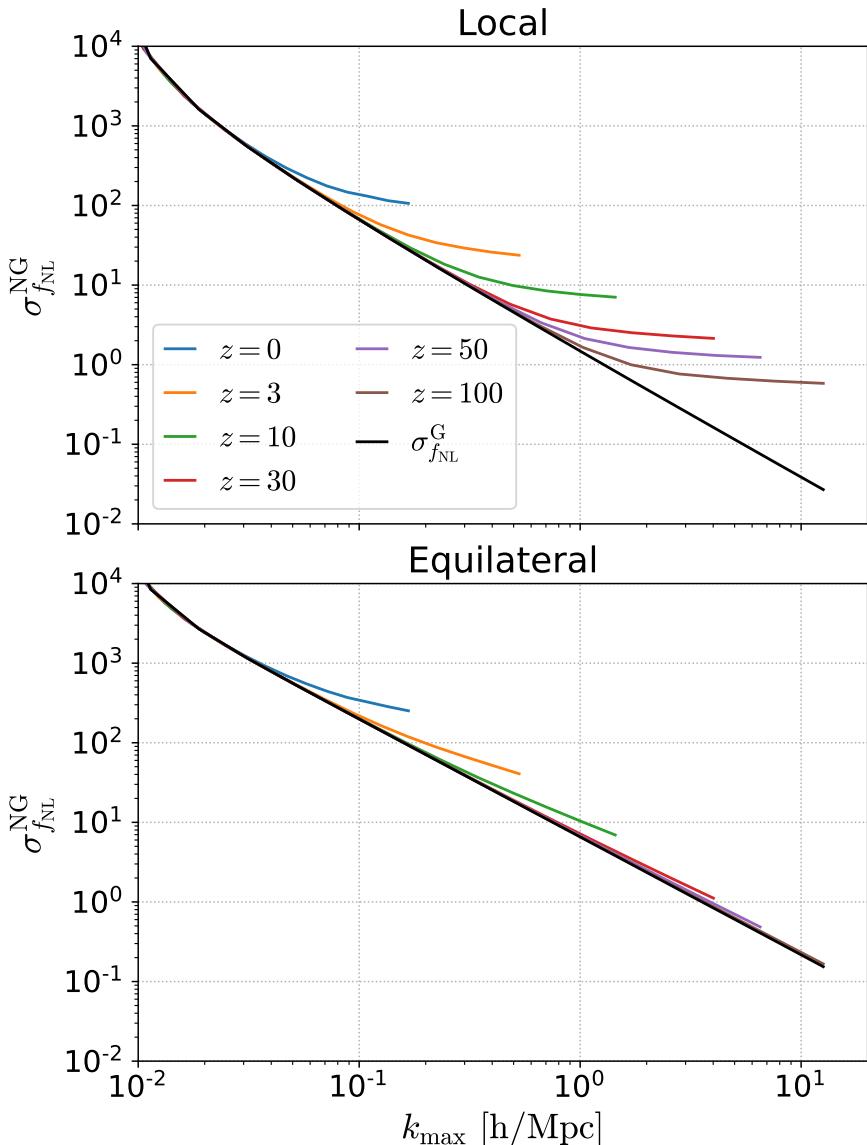
Figure 3.2: *The estimated uncertainty on f_{NL} as a function of k_{max} in the matter field at different redshifts, when including non-Gaussian covariance. The volume of the survey is taken to be $1\,(\mathrm{Gpc/h})^3$. Each redshift is shown up to the corresponding non-linear scale k_{NL}. Solid black is the (redshift independent) uncertainty assuming only Gaussian covariance.*

sian covariance. To compute the non-Gaussian covariance, we use the model of Eq. (3.16). We then proceed to compute the Fisher matrix, which is estimated in the thin bins form of Eq. (3.14). We marginalize over all the 7 nuisance parameters entering the bispectrum. For more details, we refer to Appendix 3.E.

Figure 3.3 shows the ratio of the estimated uncertainty computed using a non-Gaussian covariance over a Gaussian approximation for the local and equilateral type non-Gaussianities as a function of redshift for a PUMA-like experiment. We compute the uncertainty for two different values of k_{max}, corresponding to $0.5\,k_{\mathrm{NL}}$ (dashed lines) and $0.75\,k_{\mathrm{NL}}$ (solid lines), as we expect Eq. (3.6) to be less accurate for tracers. Our results show that even for a more conservative choice of $k_{\mathrm{max}} = 0.5\,k_{\mathrm{NL}}$, the effect is significant. The increase in uncertainty ranges from a factor of 2 to a factor of 5 for local type non-Gaussianity. We conclude that previous forecasts on constraining f_{NL} at high-redshift are too optimistic [20, 22–25] and non-Gaussian covariance will have to be considered in order to produce more realistic forecasts. A similar estimation for a generic biased tracer was performed in [46] up to $z = 10$ and shows qualitative agreement with Figure 3.3.
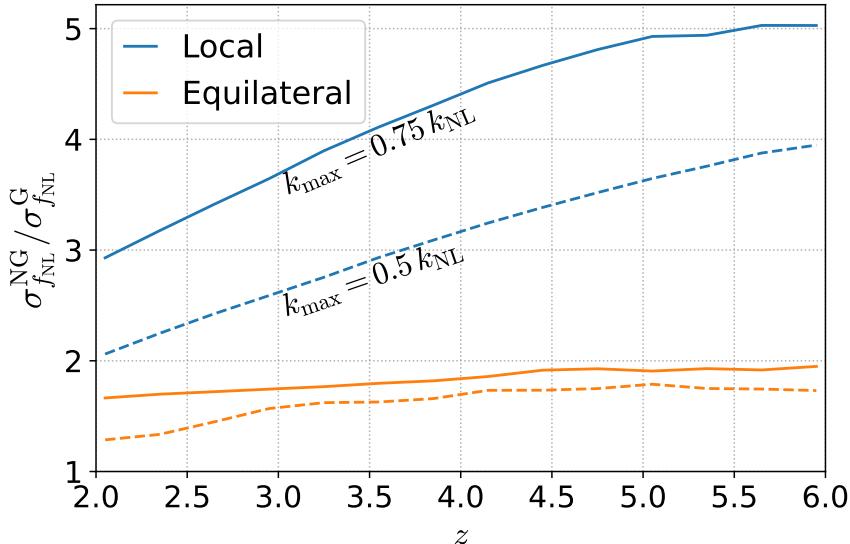
Figure 3.3: *Estimate of the relative increase in error on the non-Gaussian amplitude f_{NL} due to non-Gaussian covariance of the hydrogen bispectrum, as a function of redshift for a PUMA-like experiment when marginalizing over the 7 additional parameters of the hydrogen bispectrum. We show the results for $k_{\mathrm{max}} = 0.75\,k_{\mathrm{NL}}$ (solid lines) and $k_{\mathrm{max}} = 0.5\,k_{\mathrm{NL}}$ (dashed lines).*

3.4   Discussion and Conclusions

We studied the impact of non-Gaussian terms in the covariance on measurements of cosmological correlators. Specifically, we aim to quantify the effect on the estimated uncertainty of the primordial non-Gaussian amplitude f_{NL} when using the bispectrum at high redshift as an observable. Because off-diagonal components are small compared to the diagonal, most studies have typically neglected this covariance. We showed that, when looking at the information content, there is a significant impact on the constraining power on primordial non-Gaussianity due to this non-Gaussian mode coupling, even at high redshifts and well below the non-linear scale as defined in Eq. (3.6).

We have first computed the effect of non-Gaussian covariance on $\sigma_{f_{NL}}$ using the matter bispectrum in real space and then performed a more realistic estimation using the hydrogen bispectrum as measured from a PUMA-like experiment. This proposed 21-cm intensity mapping experiment has the potential to constrain primordial bispectra to reach beyond constraints set by the CMB. Yet, our analysis shows that not accounting for the full covariance can overestimate the constraining power of the hydrogen bispectrum measured by PUMA up to a factor of 5 for local type non-Gaussianity and 2 for equilateral. For local-type non-Gaussianity, the primary observable is actually the tracer power spectrum, thanks to the so-called scale-dependent bias, which we do not include in our analysis. Nevertheless, our results imply that combining it with the bispectrum does not help as much as it is expected to when considering a Gaussian covariance only. Moreover, they motivate including the bispectrum-power spectrum cross-covariance in the joint analysis, which is also a non-Gaussian contribution [18]. Overall, our result suggests we should reconsider some of the existing forecasts and make sure the projected numbers are not overly optimistic for future high redshift surveys such as PUMA, MegaMapper [47] and the Maunakea Spectroscopic Explorer [48].

Future constraints on primordial non-Gaussianities will depend on our ability to extract information from large-scale structures. Intuitively, the main obstacle to constraining primordial spectra is set by the non-linear scale which estimates when loop-corrections become important. Here we show that for the Fisher information on f_{NL} it is important to account for non-Gaussian bispectrum covariance, even for modes that are still considered linear. The results are comparable to what was found for measurements of the CMB bispectrum, where lensing-induced off-diagonal covariance is the main limitation as we start to measure smaller scales and increase the number of accessible modes.

For the CMB, it was shown that the lensing-induced covariance can be accounted for by delensing the data before applying the standard estimators [38]. The analogy here would be to "degravitate" the data, a technique that is well established in studies of the Baryon Acoustic Oscillations in galaxy surveys [49]. In fact, this approach has been shown to improve constraints on primordial non-Gaussianity from the bispectrum [50]. It might be possible to explore this option at high redshifts, where the physics is still perturbatively tractable. At lower redshifts, however, it likely suggests that existing estimators are sub-optimal or that we have adopted inefficient summary statistics that need to be revisited. Similar conclusions were drawn in Ref. [51]. Applying reconstruction methods [52] or using simulations (e.g. through simulation-based inference [53]) [54–58], both active fields of investigation, will certainly help to establish to what degree we have to modify our analysis tools in search for signs of primordial non-Gaussianity.

The code used to produce the results in this chapter is publicly available [6].

---

[6] https://github.com/tsfloss/pyNG

3.A   Primordial Bispectra

The local, equilateral, and orthogonal bispectrum templates are given by:

$$
B_\zeta^{\text{loc}}(k_1, k_2, k_3) = \frac{6}{5} f_{\text{NL}}^{\text{loc}} \left(P_{\zeta,1} P_{\zeta,2} + P_{\zeta,1} P_{\zeta,3} + P_{\zeta,2} P_{\zeta,3} \right)
\tag{3.17}
$$

$$
B_\zeta^{\text{equil}}(k_1, k_2, k_3) = \frac{18}{5} f_{\text{NL}}^{\text{equil}} \Big(-(P_{\zeta,1} P_{\zeta,2} + 2 \text{ perms.}) - 2 P_{\zeta,1}^{2/3} P_{\zeta,2}^{2/3} P_{\zeta,3}^{2/3} +
$$
$$
(P_{\zeta,1}^{1/3} P_{\zeta,2}^{2/3} P_{\zeta,3} + 5 \text{ perms.}) \Big)
\tag{3.18}
$$

$$
B_\zeta^{\text{ortho}}(k_1, k_2, k_3) = \frac{18}{5} f_{\text{NL}}^{\text{ortho}} \Big(-3(P_{\zeta,1} P_{\zeta,2} + 2 \text{ perms.}) - 8 P_{\zeta,1}^{2/3} P_{\zeta,2}^{2/3} P_{\zeta,3}^{2/3} +
$$
$$
3(P_{\zeta,1}^{1/3} P_{\zeta,2}^{2/3} P_{\zeta,3} + 5 \text{ perms.}) \Big) ,
\tag{3.19}
$$

where we introduced the shorthand notation $P_{\zeta,1} = P_\zeta(k_1)$. Note that although Eq. (3.2) demonstrates that the primordial bispectrum can in principle depend on the full three-momenta, for the above primordial shapes there is no angular dependence and they only depend on the magnitudes of the triangle's momenta (i.e. the shape of the triangle).

3.B   Standard Perturbation Theory at Tree Level

Within the perturbative regime, gravitational interactions can be treated within the framework of Standard Perturbation Theory (SPT). For an extensive review, we refer to e.g. [33]. Here we will present only the results relevant to our work. Since we require the gravitational trispectrum, we expand the density field to third order:

$$
\delta_{\mathbf{k}}(z) = \delta_{\mathbf{k}}^{(1)}(z) + \delta_{\mathbf{k}}^{(2)}(z) + \delta_{\mathbf{k}}^{(3)},
\tag{3.20}
$$

where the superscript denotes the order of the perturbation. Solving the evolution equations order by order in perturbations one finds (dropping the explicit time-dependence):

$$
\delta_{\mathbf{k}}^{(2)} = \int_{\mathbf{q}} F_2(\mathbf{q}, \mathbf{k} - \mathbf{q}) \delta_{\mathbf{q}}^{(1)} \delta_{\mathbf{k}-\mathbf{q}}^{(1)},
$$
$$
\delta_{\mathbf{k}}^{(3)} = \int_{\mathbf{q}_1, \mathbf{q}_2} F_3(\mathbf{q}_1, \mathbf{q}_2, \mathbf{k} - \mathbf{q}_{12}) \delta_{\mathbf{q}_1}^{(1)} \delta_{\mathbf{q}_2}^{(1)} \delta_{\mathbf{k}-\mathbf{q}_{12}}^{(1)}.
\tag{3.21}
$$

These higher-order perturbations will induce the gravitational correlations. At tree level then, the bispectrum of the density field due to gravitational interactions, is

found to be:

$$
\begin{aligned}
B_\delta^{\mathrm{snG}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) &= \langle \delta_{\mathbf{k}_1}^{(1)} \delta_{\mathbf{k}_2}^{(1)} \delta_{\mathbf{k}_3}^{(2)} \rangle + 2 \text{ perms.} \\
&= 2 F_2(\mathbf{k}_1, \mathbf{k}_2) P_\delta^L(k_1) P_\delta^L(k_2) + 2 \text{ perms.},
\end{aligned}
$$
(3.22)

where the linear power spectrum is defined as:

$$
\begin{aligned}
\langle \delta_{\mathbf{k}_1}^{(1)} \delta_{\mathbf{k}_2}^{(1)} \rangle &= (2\pi)^3 \delta_D(\mathbf{k}_{12}) P_\delta^L(k_1) \\
&= (2\pi)^3 \delta_D(\mathbf{k}_{12}) \mathcal{M}(k_1, z) P_\zeta(k_1).
\end{aligned}
$$
(3.23)

Here $\mathcal{M}$ is the linear transfer function. The tree-level trispectrum consists of two contributions:

$$
T_\delta^{\mathrm{snG}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) = T_\delta^{1122}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) + T_\delta^{1113}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4)
$$
(3.24)

with the two contributions given by:

$$
\begin{aligned}
T_\delta^{1113}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) &= \langle \delta_{\mathbf{k}_1}^{(1)} \delta_{\mathbf{k}_2}^{(1)} \delta_{\mathbf{k}_3}^{(1)} \delta_{\mathbf{k}_4}^{(3)} \rangle + 3 \text{ perms.} \\
&= 6 F_3(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \prod_{i=1}^{3} P_\delta^L(k_i) + 3 \text{ perms.} \\
T_\delta^{1122}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) &= \langle \delta_{\mathbf{k}_1}^{(1)} \delta_{\mathbf{k}_2}^{(1)} \delta_{\mathbf{k}_3}^{(2)} \delta_{\mathbf{k}_4}^{(2)} \rangle + 5 \text{ perms.} \\
&= 4 \left[F_2(-\mathbf{k}_1, \mathbf{k}_{13}) P_\delta^L(k_{13}) + F_2(-\mathbf{k}_1, \mathbf{k}_{14}) P_\delta^L(k_{14}) \right] \\
&\quad \times P_\delta^L(k_1) P_\delta^L(k_2) + 5 \text{ perms.}
\end{aligned}
$$
(3.25)

The kernels F_2, F_3 determine how modes of different wavelengths are coupled by gravity. For their explicit form, see e.g. [33]. Beyond tree-level, the higher-order perturbations will also induce a correction to the power spectrum, known as the 1-loop power spectrum:

$$
\langle \delta_{\mathbf{k}_1}^{(2)} \delta_{\mathbf{k}_2}^{(2)} \rangle = (2\pi)^3 \delta_D(\mathbf{k}_{12}) P_\delta^{1\text{-loop}}(k_1).
$$
(3.26)

3.C Details on simulations

In order to verify our computations of non-Gaussian covariance, we compare our results to simulations. For redshifts $z = 0$ and 3 we use the QUIJOTE simulation suite [28]. QUIJOTE consists of 15000 N-body simulations using a fiducial cosmology, enough to obtain an accurate covariance matrix up to the scales of interest in this chapter. The simulations consist of 512^3 particles in a box with sides 1000 Mpc/h,

setting the fundamental mode to $k_F = (2\pi/1000) \approx 0.0063$ h/Mpc. To construct the covariance matrix, we use the power spectrum and bispectrum measurements as provided in the suite, which uses a binning of $\Delta k = 3\,k_F$, $k_{\min} = \frac{3}{2}k_F$ and an interpolation grid of size 360^3. For redshift $z = 10$ we use ~ 4000 realizations of initial conditions (ICs) generated using third-order Lagrangian perturbation theory (3LPT) with the `Monofonic` code [36]. Since at higher redshifts, the power spectrum and bispectrum are smaller, shot-noise becomes increasingly dominant. The $z = 10$ realizations are therefore generated with 512^3 particles in a box of 250 Mpc/h, setting the fundamental mode to $k_F = (2\pi/250) \approx 0.025$ h/Mpc. Measurements of the power spectrum and bispectrum are done using the codes `Pylians` [7] and `PySpectrum` [8] respectively, with the same settings as used for the QUIJOTE measurements. Having measured the power spectrum and bispectrum from many realizations of the simulation, we obtain the covariance matrix through Eq. (3.8). When inverting the covariance matrix from simulations, we include the Hartlap factor to unbias the numerical matrix [59].

Evaluation of bin-averaged primordial bispectra

In order to compute the Fisher information for primordial bispectra from the simulations, we need the bin-averaged derivatives of the theoretical bispectrum with respect to f_{NL} that appear in Eq. (4.14):

$$
\begin{aligned}
\frac{\partial \hat{B}(k_1, k_2, k_3)}{\partial f_{\mathrm{NL}}} &= \frac{1}{V_{123}} \int_{k_1} d^3\mathbf{q}_1 \int_{k_2} d^3\mathbf{q}_2 \int_{k_3} d^3\mathbf{q}_3 \frac{\partial B(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3)}{\partial f_{\mathrm{NL}}} \\
&= \frac{1}{V_{123}} \int_{k_1} d^3\mathbf{q}_1 \int_{k_2} d^3\mathbf{q}_2 \int_{k_3} d^3\mathbf{q}_3 \left(\prod_{i=1}^{3} \mathcal{M}(q_i, z) \right) B_\zeta(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3)|_{f_{\mathrm{NL}}=1},
\end{aligned}
\tag{3.27}
$$

where the hat denotes a bin-averaged quantity and the volume of the bin is given by

$$
V_{123} = \int_{k_1} d^3\mathbf{q}_1 \int_{k_2} d^3\mathbf{q}_2 \int_{k_3} d^3\mathbf{q}_3
\tag{3.28}
$$

and the integrals denote a binning similar to that of the simulation measurements, i.e. over spherical shells with centers k_i and width $[k_i - 3k_F/2, k_i + 3k_F/2]$.

Orthogonal shape

When computing the bin-averaged orthogonal shape bispectrum, one realizes that it becomes negative for certain triangle configurations. For the coarse binning ($\Delta k =$

---

[7]`https://github.com/franciscovillaescusa/Pylians`
[8]`https://github.com/changhoonhahn/pySpectrum`
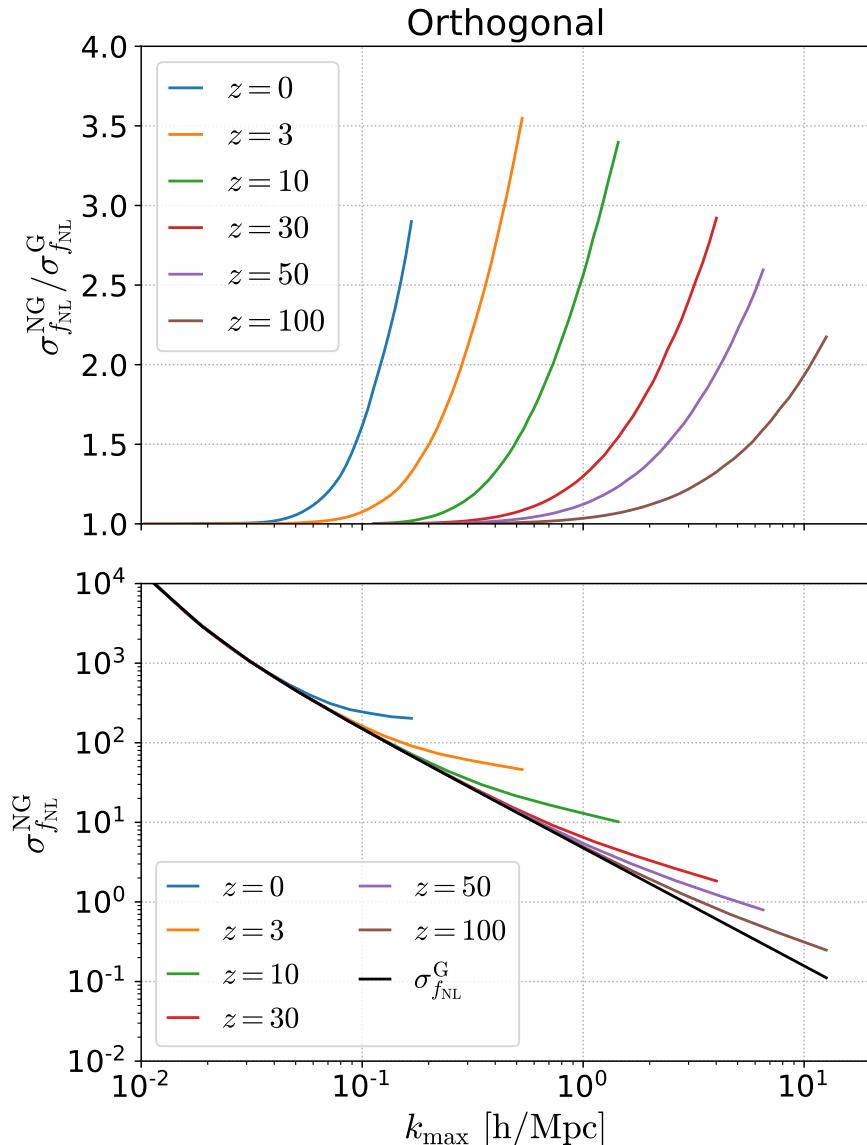
Figure 3.4: *Similar plots as Figure 3.1 and Figure 3.2 but now for orthogonal non-Gaussianity.*

$3k_F$) of the simulation data we use, the bin-averaged bispectrum in Eq. (3.27) suffers from cancellations within the bin, removing part of the signal. Since remeasuring the bispectrum in QUIJOTE simulations with finer binning goes beyond the scope of this

chapter, we decided to omit these results. Nonetheless, our predictions in the thin bin approximation are presented in Figure 3.4 and for $z = 0$ agree well with the simulation results presented in [51].

3.D   Details on the Fisher matrix in the thin bins limit

To produce our theoretical predictions of the loss of constraining power due to non-Gaussian covariance, we take the continuous limit of Eq. (4.14) while approximating the inverse covariance matrix.

As explained in the main text, this allows us to probe a wide range of scales at low computational cost. The main complication for Eq. (3.13) is to explicitly compute the inverse covariance as a function of triangle configurations. To this end, we expand the inverse using a Neumann series:

$$C^{-1} \;=\; \sum_{n=0}^{\infty} \left(-C_{\mathrm{G}}^{-1} C_{\mathrm{nG}}\right)^n C_{\mathrm{G}}^{-1}, \tag{3.29}$$

such that we never have to invert the non-Gaussian covariance matrix that contains off-diagonal terms. Hence the Fisher matrix becomes the sum of infinitely many terms:

$$F_{ab} = \sum_{n=0}^{\infty} \sum_{TT'} \frac{\partial B_T}{\partial p_a} \left(\left(-C_{\mathrm{G}}^{-1} C_{\mathrm{nG}}\right)^n C_{\mathrm{G}}^{-1}\right)_{TT'} \frac{\partial B_T}{\partial p_b}. \tag{3.30}$$

Since the terms in this sum are increasingly complicated to compute (in the continuous limit, every matrix multiplication becomes an integral over triangle configurations), we choose to approximate the expansion using:

$$
\begin{aligned}
F_{ab} &= \sum_{TT'} \frac{\partial B_T}{\partial p_a} \left(C_{\mathrm{G}}^{-1}\right)_{TT'} \frac{\partial B_T}{\partial p_b} + \sum_{n=1}^{\infty} \frac{\left(\sum_{TT'} \frac{\partial B_T}{\partial p_a} \left(-C_{\mathrm{G}}^{-1} C_{\mathrm{nG}} C_{\mathrm{G}}^{-1}\right)_{TT'} \frac{\partial B_T}{\partial p_b}\right)^n}{\left(\sum_{TT'} \frac{\partial B_T}{\partial p_a} \left(C_{\mathrm{G}}^{-1}\right)_{TT'} \frac{\partial B_T}{\partial p_b}\right)^{n-1}} \\
&= F_{ab}^{G} + \sum_{n=1}^{\infty} (-1)^n \frac{\left(\delta F_{ab}^{\mathrm{nG}}\right)^n}{\left(F_{ab}^{\mathrm{G}}\right)^{n-1}}
\end{aligned}
\tag{3.31}
$$

which can be recognized as the expansion of Eq. (3.14). The approximation of Eq. (3.31) seems to work reasonably well in the range of k_{max} we compare with simulation results (see Fig. 3.1 in the main text). On the other hand, our key result is that these non-Gaussian terms are actually important, therefore we expect our approximation to break down. This motivates further work in defining a proper estimator for primordial non-Gaussianity in the presence of non-Gaussian covariance terms.

3.E Details on the PUMA analysis

We perform the analysis of the impact of non-Gaussian covariance on the PUMA survey along the lines of Ref. [23]. We use a binning of $\Delta z = 0.1$ between $2 < z < 6$. The largest available scale is set by the volume $V_s(z)$ of the redshift bin through $k_{\min}(z) = k_F(z) = 2\pi/L(z)$, where $L(z) = V_s(z)^{1/3}$, $V_s(z) = \frac{4\pi}{3}(r(z + \Delta z)^3 - r(z - \Delta z)^3)$ and $r(z)$ is the comoving distance to redshift z in units of Mpc/h.

Foregrounds

The largest scale is set by a foreground cut in the line-of-sight direction ($k_{\parallel,\min} = 0.01$ h/Mpc), removing much of the dependence on the choice of redshift binning. We limit ourselves to linear scales by choosing $k_{\max}(z)$ to be $0.5\,k_{\mathrm{NL}}(z)$ or $0.75\,k_{\mathrm{NL}}(z)$. We model the foreground wedge by excluding all modes for which

$$k_{\parallel} < \frac{r(z)H(z)}{c(1+z)}\sin\left(0.66 N_w \theta_{\mathrm{FOV}}(z)\right) \times k_{\perp}. \tag{3.32}$$

Here $H(z)$ is the Hubble parameter, $\theta_{\mathrm{FOV}}(z) = \lambda_{21}(z)/D_{\mathrm{eff}}$, $\lambda_{21}(z)$ is the redshifted 21-cm wavelength in meters and $D_{\mathrm{eff}} = (\sqrt{0.7} \times 6 \text{ meters})$ is the effective dish size of the PUMA setup. N_w determines the severity of the foreground wedge. We apply a pessimistic wedge cut of $N_w = 3$ and a foreground cut $k_{\parallel,\min} = 0.01$ h/Mpc, in order to show that the loss of constraining power as due to non-Gaussian covariance persists in such a setup. Finally, the largest and smallest accessible scales in the perpendicular direction are set by:

$$k_{\perp,\max}(z) = \frac{2\pi D_{\max}}{\lambda_{21}(z)r(z)}, \qquad k_{\perp,\min}(z) = \frac{2\pi}{r(z)\theta_{\mathrm{FOV}}(z)}, \tag{3.33}$$

where $D_{\max} = 700$ meters is the largest baseline of the PUMA setup. We show all the relevant scales together in Figure 3.5.

The hydrogen power spectrum and bispectrum

The calculation of hydrogen correlation functions in redshift-space is significantly more involved than the matter field case from the modeling point of view. Besides non-linearities in the matter field, we need to account for the biased relation between the hydrogen and matter distributions, redshift space distortions, and stochasticity introduced by the discreteness effects and Poisson noise. A complete explanation of these modeling efforts can be found in [23]. Here we quote the hydrogen power

spectrum and bispectrum for reference, defined as

$$P_{\mathrm{HI}}(z, \mathbf{k}) = P_N(z, \mathbf{k}) + T_b(z)^2 D_{\mathrm{FOG}}^P(z, \mathbf{k}) \left[Z_1(z, \mathbf{k})^2 P_\delta^L(z, k) + P_\varepsilon(z) \right],$$

$$(3.34)$$

and

$$
\begin{aligned}
B_{\mathrm{HI}}(z, \mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = {} & T_b(z)^3 D_{\mathrm{FOG}}^B(z, \mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \times \\
& \left[\prod_{i=1}^3 Z_1(z, \mathbf{k}_i) B_\delta^{\mathrm{pnG}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \right. \\
& \quad + 2 Z_1(z, \mathbf{k}_1) Z_1(z, \mathbf{k}_2) Z_2(z, \mathbf{k}_1, \mathbf{k}_2) P_\delta^L(z, \mathbf{k}_1) P_\delta^L(z, \mathbf{k}_2) \\
& \left. \quad + 2 \text{ perm.} \right] \\
& + P_{\varepsilon\varepsilon\delta}(z) \left[\sum_{i=1}^3 Z_1(z, \mathbf{k}_i) P_\delta^L(z, k_i) \right] \\
& + B_\varepsilon(z),
\end{aligned}
$$

$$(3.35)$$

where P_N is the instrumental noise, $T_b(z)$ is the brightness temperature of the 21-cm signal at a given redshift, P_ε, $P_{\varepsilon\varepsilon\delta}$, B_ε are stochastic noise contributions, Z_1, Z_2 are the first and second order redshift space kernels, B_δ^{pnG} is the primordial contribution to the matter bispectrum as used in Eq. (3.27) and D_{FOG}^B models the Finger-Of-God damping effect. For the explicit expressions of these quantities, we refer to [23]. The redshift space kernels contain bias parameters $\{b_1, b_2, b_{s^2}, b_\Psi, b_{\Psi\delta}\}$ as well as the linear growth rate f due to redshift space distortions RSDs. The scale-dependent biases $\{b_\Psi, b_{\Psi\delta}\}$ can be modelled in terms of $\{b_1, b_2, f_{\mathrm{NL}}\}$ (though see [60] for a study of this approximation). This means primordial non-Gaussianity enters not only through B_δ^{pnG} but also through the terms involving Z_1, Z_2, that contain the scale-dependent biases. Finally, the FOG factor is modeled using the velocity dispersion σ_v. In this chapter we are only interested in signal-to-noise for f_{NL} coming from the hydrogen bispectrum, hence the total number of parameters including the stochastic noise contributions equals 8:

$$\mathbf{p} = \{f_{\mathrm{NL}}, b_1, b_2, b_{s^2}, f, \sigma_v, P_{\varepsilon\varepsilon\delta}, B_\varepsilon\} \tag{3.36}$$

We calculate the Fisher matrix of the 8 parameters that enter the hydrogen bispectrum, with and without non-Gaussian covariance at each redshift bin using the weighted estimator approach described above [?]. Contrary to [23] we do not account for theoretical errors on the bias parameters in our analysis, which adds additional covariance (including off-diagonal) to account for uncertainties in the bias

model along the lines of Ref. [61]. Once we have the Gaussian and non-Gaussian Fisher matrices, we marginalize over the 7 nuisance parameters by inverting the Fisher matrix at every redshift. The marginalised uncertainty for f_{NL} is then given by:

$$\sigma_{f_{\mathrm{NL}}}(z) = \left(F^{-1}(z)\right)^{1/2}_{f_{\mathrm{NL}} f_{\mathrm{NL}}} \tag{3.37}$$

The ratio of the estimated uncertainty including non-Gaussian covariance over Gaussian covariance is shown in Figure 3.3.
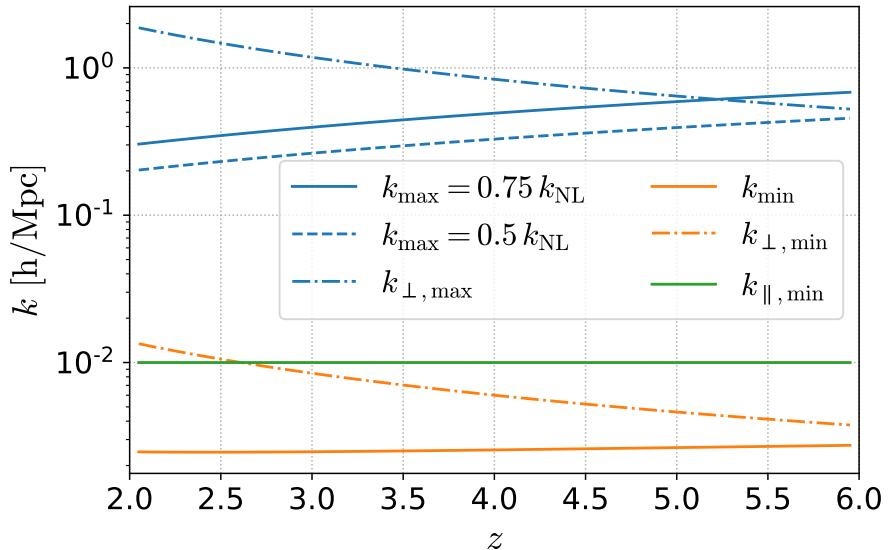
Figure 3.5: *Smallest and largest accessible scales as a function of redshift for the PUMA survey. The smallest overall scale k_{max} is determined by the non-linear scale. The largest overall scale k_{min} is set by the volume of the redshift bin and hence depends on the choice of binning. The smallest and largest scales in the perpendicular direction, $k_{\perp,\mathrm{max}}$ and $k_{\perp,\mathrm{min}}$ are set by properties of the experiment. Finally, we apply a foreground cut $k_{\parallel,\mathrm{min}} = 0.01 \ h/Mpc$ in the line-of-sight direction, that effectively replaces k_{min}, removing most dependence on the choice of binning.*

Bibliography

[1] A. H. Guth, "The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems," *Phys. Rev. D* **23**, pp. 347–356, 1981.

[2] A. D. Linde, "A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems," *Phys. Lett. B* **108**, pp. 389–393, 1982.

[3] A. Albrecht and P. J. Steinhardt, "Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking," *Phys. Rev. Lett.* **48**, pp. 1220–1223, 1982.

[4] Y. Akrami *et al.*, "Planck 2018 results. X. Constraints on inflation," 2018.

[5] P. D. Meerburg *et al.*, "Primordial Non-Gaussianity," 3 2019.

[6] A. Achúcarro *et al.*, "Inflation: Theory and Observations," 3 2022.

[7] J. M. Maldacena, "Non-Gaussian features of primordial fluctuations in single field inflationary models," *JHEP* **05**, p. 013, 2003.

[8] P. Creminelli and M. Zaldarriaga, "Single field consistency relation for the 3-point function," *JCAP* **10**, p. 006, 2004.

[9] N. Arkani-Hamed and J. Maldacena, "Cosmological Collider Physics," 2015.

[10] R. Takahashi, N. Yoshida, M. Takada, T. Matsubara, N. Sugiyama, I. Kayo, T. Nishimichi, S. Saito, and A. Taruya, "Non-Gaussian Error Contribution to Likelihood Analysis of the Matter Power Spectrum," *Astrophys. J.* **726**, p. 7, 2011.

[11] K. C. Chan and L. Blot, "Assessment of the Information Content of the Power Spectrum and Bispectrum," *Phys. Rev. D* **96**(2), p. 023528, 2017.

[12] K. C. Chan, A. Moradinezhad Dizgah, and J. Noreña, "Bispectrum Supersample Covariance," *Phys. Rev. D* **97**(4), p. 043532, 2018.

[13] D. Wadekar and R. Scoccimarro, "Galaxy power spectrum multipoles covariance in perturbation theory," *Phys. Rev. D* **102**(12), p. 123517, 2020.

[14] A. Barreira, "The squeezed matter bispectrum covariance with responses," *JCAP* **03**, p. 008, 2019.

[15] D. Gualdi and L. Verde, "Galaxy redshift-space bispectrum: the Importance of Being Anisotropic," *JCAP* **06**, p. 041, 2020.

[16] A. Oddo, F. Rizzo, E. Sefusatti, C. Porciani, and P. Monaco, "Cosmological parameters from the likelihood analysis of the galaxy power spectrum and bispectrum in real space," *JCAP* **11**, p. 038, 2021.

[17] A. Barreira, "Predictions for local PNG bias in the galaxy power spectrum and bispectrum and the consequences for f $_{NL}$ constraints," *JCAP* **01**(01), p. 033, 2022.

[18] M. Biagetti, L. Castiblanco, J. Noreña, and E. Sefusatti, "The Covariance of Squeezed Bispectrum Configurations," 11 2021.

[19] F. Rizzo, C. Moretti, K. Pardede, A. Eggemeier, A. Oddo, E. Sefusatti, C. Porciani, and P. Monaco, "The Halo Bispectrum Multipoles in Redshift Space," 4 2022.

[20] J. B. Muñoz, Y. Ali-Haïmoud, and M. Kamionkowski, "Primordial non-gaussianity from the bispectrum of 21-cm fluctuations in the dark ages," *Phys. Rev. D* **92**(8), p. 083508, 2015.

[21] X. Chen, P. D. Meerburg, and M. Münchmeyer, "The Future of Primordial Features with 21 cm Tomography," *JCAP* **09**, p. 023, 2016.

[22] P. D. Meerburg, M. Münchmeyer, J. B. Muñoz, and X. Chen, "Prospects for Cosmological Collider Physics," *JCAP* **03**, p. 050, 2017.

[23] D. Karagiannis, A. Slosar, and M. Liguori, "Forecasts on Primordial non-Gaussianity from 21 cm Intensity Mapping experiments," *JCAP* **11**, p. 052, 2020.

[24] T. Flöss, T. de Wild, P. D. Meerburg, and L. V. E. Koopmans, "The Dark Ages' 21-cm Trispectrum," 1 2022.

[25] D. Yamauchi, "Signature of primordial non-Gaussianity on 21-cm power spectrum from dark ages," 3 2022.

[26] D. Karagiannis, R. Maartens, and L. Randrianjanahary, "Cosmological constraints from the power spectrum and bispectrum of 21cm intensity maps," 6 2022.

[27] R. Ansari *et al.*, "Inflation and Early Dark Energy with a Stage II Hydrogen Intensity Mapping experiment," 10 2018.

[28] F. Villaescusa-Navarro *et al.*, "The Quijote simulations," *Astrophys. J. Suppl.* **250**(1), p. 2, 2020.

[29] N. Aghanim *et al.*, "Planck 2018 results. VI. Cosmological parameters," *Astron. Astrophys.* **641**, p. A6, 2020. [Erratum: Astron.Astrophys. 652, C4 (2021)].

[30] P. A. R. Ade *et al.*, "Planck 2015 results. XIII. Cosmological parameters," *Astron. Astrophys.* **594**, p. A13, 2016.

[31] H. Lee, D. Baumann, and G. L. Pimentel, "Non-Gaussianity as a Particle Detector," *JHEP* **12**, p. 040, 2016.

[32] C. Cheung, P. Creminelli, A. L. Fitzpatrick, J. Kaplan, and L. Senatore, "The Effective Field Theory of Inflation," *JHEP* **03**, p. 014, 2008.

[33] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro, "Large scale structure of the universe and cosmological perturbation theory," *Phys. Rept.* **367**, pp. 1–248, 2002.

3

[34] J. Tomlinson and D. Jeong, "Perturbation Theory vs Simulation: Quasi-linear Scale, Binning Effect, and Visualization of Bispectrum," 4 2022.

[35] R. E. Angulo and O. Hahn, "Large-scale dark matter simulations," 12 2021.

[36] M. Michaux, O. Hahn, C. Rampf, and R. E. Angulo, "Accurate initial conditions for cosmological N-body simulations: Minimizing truncation and discreteness errors," *Mon. Not. Roy. Astron. Soc.* **500**(1), pp. 663–683, 2020.

[37] R. Scoccimarro, S. Colombi, J. N. Fry, J. A. Frieman, E. Hivon, and A. Melott, "Nonlinear evolution of the bispectrum of cosmological perturbations," *Astrophys. J.* **496**, p. 586, 1998.

[38] W. R. Coulton, P. D. Meerburg, D. G. Baker, S. Hotinli, A. J. Duivenvoorden, and A. van Engelen, "Minimizing gravitational lensing contributions to the primordial bispectrum covariance," *Phys. Rev. D* **101**(12), p. 123504, 2020.

[39] V. Desjacques, D. Jeong, and F. Schmidt, "Large-Scale Galaxy Bias," *Phys. Rept.* **733**, pp. 1–193, 2018.

[40] N. Dalal, O. Dore, D. Huterer, and A. Shirokov, "The imprints of primordial non-gaussianities on large-scale structure: scale dependent bias and abundance of virialized objects," *Phys. Rev. D* **77**, p. 123514, 2008.

[41] S. Matarrese and L. Verde, "The effect of primordial non-Gaussianity on halo bias," *Astrophys. J. Lett.* **677**, pp. L77–L80, 2008.

[42] A. Slosar, C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan, "Constraints on local primordial non-Gaussianity from large scale structure," *JCAP* **08**, p. 031, 2008.

[43] M. Biagetti, "The Hunt for Primordial Interactions in the Large Scale Structures of the Universe," *Galaxies* **7**(3), p. 71, 2019.

[44] N. Sailer, E. Castorina, S. Ferraro, and M. White, "Cosmology at high redshift — a probe of fundamental physics," *JCAP* **12**(12), p. 049, 2021.

[45] E. Castorina and F. Villaescusa-Navarro, "On the spatial distribution of neutral hydrogen in the Universe: bias and shot-noise of the HI power spectrum," *Mon. Not. Roy. Astron. Soc.* **471**(2), pp. 1788–1796, 2017.

[46] R. de Putter, "Primordial physics from large-scale structure beyond the power spectrum," 2 2018.

[47] D. J. Schlegel *et al.*, "Astro2020 APC White Paper: The MegaMapper: a $z > 2$ Spectroscopic Instrument for the Study of Inflation and Dark Energy," 7 2019.

[48] C. Babusiaux *et al.*, "The Detailed Science Case for the Maunakea Spectroscopic Explorer, 2019 edition," 4 2019.

[49] D. J. Eisenstein, H.-j. Seo, E. Sirko, and D. Spergel, "Improving Cosmological Distance Measurements by Reconstruction of the Baryon Acoustic Peak," *Astrophys. J.* **664**, pp. 675–679, 2007.

[50] M. Shirasaki, N. S. Sugiyama, R. Takahashi, and F.-S. Kitaura, "Constraining primordial non-Gaussianity with postreconstructed galaxy bispectrum in redshift space," *Phys. Rev. D* **103**(2), p. 023506, 2021.

[51] W. R. Coulton, F. Villaescusa-Navarro, D. Jamieson, M. Baldi, G. Jung, D. Karagiannis, M. Liguori, L. Verde, and B. D. Wandelt, "Quijote-PNG: Simulations of primordial non-Gaussianity and the information content of the matter field power spectrum and bispectrum," 6 2022.

[52] F. Leclercq and B. Wandelt, "Bayesian large-scale structure inference: initial conditions and the cosmic web," *IAU Symp.* **306**, pp. 1–4, 2014.

[53] K. Cranmer, J. Brehmer, and G. Louppe, "The frontier of simulation-based inference," *Proceedings of the National Academy of Sciences* **117**(48), pp. 30055–30062, 2020.

[54] J. Alsing, B. Wandelt, and S. Feeney, "Massive optimal data compression and density estimation for scalable, likelihood-free inference in cosmology," *Mon. Not. Roy. Astron. Soc.* **477**(3), pp. 2874–2885, 2018.

[55] J. Alsing, T. Charnock, S. Feeney, and B. Wandelt, "Fast likelihood-free cosmology with neural density estimators and active learning," *Mon. Not. Roy. Astron. Soc.* **488**(3), pp. 4440–4458, 2019.

[56] N. Jeffrey and B. D. Wandelt, "Solving high-dimensional parameter inference: marginal posterior densities & Moment Networks," in *34th Conference on Neural Information Processing Systems*, 11 2020.

[57] B. K. Miller, A. Cole, G. Louppe, and C. Weniger, "Simulation-efficient marginal posterior estimation with swyft: stop wasting your precious time," 11 2020.

[58] A. Cole, B. K. Miller, S. J. Witte, M. X. Cai, M. W. Grootes, F. Nattino, and C. Weniger, "Fast and Credible Likelihood-Free Cosmology with Truncated Marginal Neural Ratio Estimation," 11 2021.

[59] J. Hartlap, P. Simon, and P. Schneider, "Why your model parameter confidences might be too optimistic: Unbiased estimation of the inverse covariance matrix," *Astron. Astrophys.* **464**, p. 399, 2007.

[60] M. Biagetti, T. Lazeyras, T. Baldauf, V. Desjacques, and F. Schmidt, "Verifying the consistency relation for the scale-dependent bias from local primordial non-Gaussianity," *Mon. Not. Roy. Astron. Soc.* **468**(3), pp. 3277–3288, 2017.

[61] T. Baldauf, M. Mirbabayi, M. Simonović, and M. Zaldarriaga, "LSS constraints with controlled theoretical uncertainties," 2 2016.

3

Chapter 4

Reconstructing Primordial non-Gaussianity

Abstract

We study the use of U-Nets in reconstructing the linear dark matter density field and its consequences for constraining cosmological parameters, in particular primordial non-Gaussianity. Our network is able to reconstruct the initial conditions of redshift $z = 0$ density fields from N-body simulations with 90% accuracy out to $k \leq 0.4$ h/Mpc, competitive with state-of-the-art reconstruction algorithms at a fraction of the computational cost. We study the information content of the reconstructed $z = 0$ density field with a Fisher analysis using the QUIJOTE simulation suite, including non-Gaussian initial conditions. Combining the pre- and post-reconstructed power spectrum and bispectrum data up to $k_{\mathrm{max}} = 0.52$ h/Mpc, we find significant improvements in all parameters. Most notably, we find a factor 3.65 (local), 3.54 (equilateral), and 2.90 (orthogonal) improvement on the marginalized errors of f_{NL} as compared to only using the pre-reconstructed data. We show that these improvements can be attributed to a combination of reduced data covariance and parameter degeneracy. The results constitute an important step towards a more optimal inference of primordial non-Gaussianity from non-linear scales.

4

4.1   Introduction

Primordial non-Gaussianity (pnG) provides an important window into the early universe and in particular the inflationary epoch [1]. It provides insight into the physics at play during this epoch, from the fields responsible for the accelerated expansion and their interactions [2] to that of auxiliary particles present during the epoch [3, 4]. So far, pnG has been most precisely constrained using the anisotropies of the cosmic microwave background (CMB) [5–10]. These anisotropies, being linearly related to the small density fluctuations sourced by inflation, provide an ideal probe for studying pnG. Observations by the Planck satellite have therefore led to the most stringent constraints on pnG to date which are consistent with Gaussian initial conditions [10]. Many models of inflation predict primordial non-Gaussianity below current upper limits. Although upcoming experiments aim to improve constraints on pnG significantly [11, 12], the CMB is ultimately limited in the number of observable modes, due to its 2D nature and diffusion (Silk) damping at small scales [13]. Advancing our understanding of the primordial universe will therefore eventually depend on our ability to constrain pnGs with additional probes.

The small anisotropies observed in the CMB evolve into the large-scale structure (LSS) of the universe. The three-dimensional comoving volume of the universe, spanning all the way from the Dark Ages until now, contains exponentially more modes than the CMB, making it a powerful probe for pnGs (see e.g. Refs. [14–18] for forecasts). The underlying dark matter density distribution can be studied using biased tracers such as galaxies and atomic and molecular spectral lines [19, 20]. Except on the largest scales, contrary to the anisotropies in the CMB, anisotropies in the LSS are intrinsically non-linear due to the non-linear nature of gravity, the formation of virialized bound structures, and astrophysical processes. This naturally results in a highly non-Gaussian density distribution. In order to improve over constraints from the CMB, we need to be able to extract information from within this non-linear regime, requiring an accurate model of the tracer. Over the past years, major progress has been made to address this theoretical challenge [21–23]. Although so far LSS surveys have put constraints on pnG that are not yet competitive with CMB [24–32], these intermediate results provide confidence and motivation to further pursue this avenue. Besides the complication of modeling the tracer field, recent work has emphasized the additional complication of mode coupling or *non-Gaussian covariance* [33], especially in the context of primordial non-Gaussianity [34–37]. Non-linear evolution couples modes of different wavelengths, resulting in a reduction of the amount of unique information contained in each mode. As a result, modeling modes deeper into the non-linear regime yields diminishing returns in terms of information, saturating parameter constraints. In order to improve constraints, ad-

ditional work needs to be done to undo this mode coupling. Put in another way, non-linear evolution spreads the information content of the initial conditions into all N-point statistics. Reconstructing the initial (linear) field thus brings information from higher-order correlation functions back to the power spectrum and bispectrum. This idea has been especially well tested in the context of reconstructing the BAO peak [38]. It has also been shown that reconstruction indeed results in improved parameter constraints [39, 40]. From this perspective, in this work we will further investigate how much of the information on cosmological parameters that is lost due to non-linear gravitational evolution can be recovered by reconstructing the initial conditions at the field level.

Given the non-linear nature of gravitational evolution, it naturally lends itself to machine learning applications. Previously, machine learning has been applied to solve the *forward* problem of emulating the outcome of N-body simulations given some initial condition [41, 42]. Moreover, it was shown that neural networks trained to emulate N-body simulations learn general properties of gravitational evolution [43]. Previous attempts at reconstructing the initial conditions using machine learning have shown promising results [44, 45].

Our work consists of two parts. First, we develop a neural network based approach to reconstructing the linear initial conditions of the late-time dark matter density field that is competitive with state-of-the-art iterative reconstruction methods (i.e. Ref. [46]) at a fraction of the computational cost. Secondly, applying our reconstruction methodology to the QUIJOTE simulation suite [35, 47], we are able to determine the improvement of marginalized parameter constraints when using pre- and post-reconstructed power spectrum and bispectrum measurements. As we will see, our reconstruction method reduces mode coupling (covariance) as well as degeneracy between cosmological parameters. This results in significant improvements of parameter constraints, in particular for primordial bispectra (up to a factor of 3.65). Such improvements are hard to realize by including more non-linear modes, both because of the complications in modeling them (however, see e.g. [37, 48] for ways around this issue) as well as the aforementioned saturation due to mode coupling.

This chapter is organized as follows. In Section 5.6 we discuss the non-Gaussianity of the dark matter density field and its implications for cosmological parameter inference. Section 4.3 discusses reconstruction algorithms followed by an extensive discussion of our neural network based reconstruction method. In Section 4.4 we study the implications of our reconstruction methodology on parameter constraints. We conclude in Section 4.5.

4.2 Non-Gaussianity of the dark matter density field

4.2.1 Primordial non-Gaussianity

In this work, we will focus on the primordial three-point correlation function of the primordial potential $\Phi(\mathbf{x})$, or its Fourier space equivalent, the primordial bispectrum:

$$\langle \Phi_{\mathbf{k}_1} \Phi_{\mathbf{k}_2} \Phi_{\mathbf{k}_3} \rangle = (2\pi)^3 \, \delta_{\mathrm{D}}^{(3)}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B_{\Phi}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \tag{4.1}$$

where spatial homogeneity forces the momentum space triangles to be closed. Different inflationary scenarios source bispectra with distinct functional dependence on the triangle configuration, often referred to as the *shape* of the bispectrum [49]. In order to look for primordial non-Gaussianity in data, bispectrum templates have been developed that cover general features of classes of inflationary theories parameterized only by their overall amplitude f_{NL}. We will study three of these templates. The first of them, known as the *local* shape is given by

$$B_{\Phi}^{\mathrm{local}}(k_1, k_2, k_3) = 2 f_{\mathrm{NL}}^{\mathrm{local}} \left(P_{\Phi}(k_1) P_{\Phi}(k_2) + 2 \, \mathrm{perms.} \right), \tag{4.2}$$

which peaks in the *squeezed limit*, i.e. when $k_1 \ll k_2 \sim k_3$. This type of non-Gaussianity is generically suppressed in inflationary models with only a single field driving the accelerated expansion, such as slow-roll inflation [50, 51]. Detection of a large bispectrum of this kind would be a strong hint toward inflationary scenarios with multiple fields. Cubic self-interactions of the inflaton, e.g. $(\partial \Phi)^3$, give rise to a bispectrum that is largest for equilateral triangle configurations, i.e. when $k_1 = k_2 = k_3$, captured by the template:

$$B_{\Phi}^{\mathrm{equil}}(k_1, k_2, k_3) = 6 f_{\mathrm{NL}}^{\mathrm{equil}} \Big(-P_{\Phi}(k_1) P_{\Phi}(k_2) + 2 \, \mathrm{perms.} - 2 \left(P_{\Phi}(k_1) P_{\Phi}(k_2) P_{\Phi}(k_3) \right)^{2/3}$$
$$+ P_{\Phi}(k_1)^{1/3} P_{\Phi}(k_2)^{2/3} P_{\Phi}(k_3) + 5 \, \mathrm{perms.} \Big). \tag{4.3}$$

These types of interactions naturally arise in the Effective Field Theory (EFT) of Inflation [52]. Another prediction of this EFT is the orthogonal shape, covered by the template [53]:

$$B_{\Phi}^{\mathrm{orth}}(k_1, k_2, k_3) = 6 f_{\mathrm{NL}}^{\mathrm{orth}} \left((1+p) \frac{\Delta_{123}}{k_1^3 k_2^3 k_3^3} - p \frac{\Gamma_{123}^3}{k_1^4 k_2^4 k_3^4} \right),$$
$$\Delta_{123} = (k_T - 2k_1)(k_T - 2k_2)(k_T - 2k_3),$$
$$\Gamma_{123} = \frac{2}{3}(k_1 k_2 + k_2 k_3 + k_3 k_1) - \frac{1}{3}(k_1^2 + k_2^2 + k_3^2),$$
$$p = \frac{27}{-21 + \frac{743}{7(20\pi^2 - 193)}} \tag{4.4}$$

The imprint of primordial non-Gaussianity on the large-scale structure of the universe is an active field of research. When studied in perturbation theory, the lowest order contribution of the primordial bispectrum to the dark matter density statistics is through the linearly evolved bispectrum:

$$B_\delta(k_1, k_2, k_3) = \mathcal{M}(k_1)\mathcal{M}(k_2)\mathcal{M}(k_3)B_\Phi(k_1, k_2, k_3), \tag{4.5}$$

where $\mathcal{M}(k)$ is the linear transfer function for dark matter fluctuations. Additionally, there are higher-order non-linear corrections to the dark matter power spectrum and bispectrum that correspond to loop corrections in the perturbation theory.

4.2.2 Gravitational non-Gaussianity

Although the universe at the time of last-scattering appears Gaussian, the late-time universe does not. Limiting ourselves to the dark matter density field underlying the large-scale structure of the universe, this non-Gaussianity is a consequence only of the non-linear gravitational evolution. Any small amount of non-Gaussianity in the initial conditions is therefore obscured by the strong gravitational non-Gaussianity. To put reliable constraints on pnGs from LSS requires an accurate model of the gravitationally induced non-Gaussianity. In addition, any uncertainties in the parameters of the model need to be marginalized over to avoid bias in the primordial parameter f_{NL} that could result in a false detection.

The highly non-linear nature of gravity makes it challenging to model using standard perturbation theory (SPT, [21]). To improve predictability, modifications to SPT, such as regPT [22] have been proposed. The Effective Field Theory of Large Scale Structure (EFTofLSS) [23] is a self-consistent expansion of the fluctuations, and limitations and corrections are well understood. EFTofLSS is currently considered the most reliable analytical perturbation expansion and it can accurately model the dark matter power spectrum and bispectrum out to scales $k \leq 0.6$ h/Mpc at the cost of having to marginalize over additional parameters [54, 55].

To push beyond these scales accessible with analytical methods, we rely on numerical N-body simulations of the dark matter field, which evolve an initial distribution of dark matter particles in a periodic box under their gravitational interaction. In principle, running many of these simulations and measuring their power spectrum and bispectrum provide an accurate representation to deep within the non-linear regime. The initial conditions can be generated assuming varying cosmological parameters, which allows studying the imprint of these parameters on the late-time statistics of the dark matter field.

4.2.3   Non-Gaussian covariance

Besides swamping any primordial signal, gravitational evolution can couple modes of different wavelengths. In the absence of this mode coupling, every Fourier mode in the density field is its own Gaussian random variable, independent of all the other modes. This remains a good approximation when there is only a weak coupling, such as a weak primordial bispectrum (e.g. in the CMB, however, see also [56]). In the presence of strong mode coupling, the information contained in the initial density field is scrambled across N-point statistics at late times. Hence, even if we are able to accurately model the power spectrum and bispectrum, the amount of information available from just the power spectrum and bispectrum is limited by this mode coupling. When inferring cosmological parameters from data, this mode coupling is captured by off-diagonal components of the covariance matrix that enters the likelihood function. These entries to the covariance matrix (together with higher N-point contributions to the diagonal) are referred to as non-Gaussian covariance. The covariance matrix can in principle be modeled analogous to the power spectrum and bispectrum signals itself, although the computation involves higher-order correlation functions (see e.g. [33, 34]), which we can model analytically only on large scales, or low k. Instead, using N-body simulations, the covariance matrix can be determined numerically by taking the measurements of many simulation realizations.

The covariance matrix plays a vital part in a Fisher analysis that forecasts the expected error bars on cosmological parameters for a particular survey (e.g. galaxies or intensity mapping). Forecasts derived by analytical calculations often assume only the Gaussian (diagonal) contribution. In Ref. [33] it was shown that this assumption is unwarranted as non-Gaussian covariance of the dark matter and halo power spectrum and bispectrum significantly reduce the signal-to-noise ratio already in what is usually considered to be the linear regime. Recently, Refs. [34] showed that the bispectrum covariance is dominated by squeezed triangle configurations, significantly affecting constraints on local primordial non-Gaussianity. This finding was extended to other shapes of non-Gaussianity using the recently released QUIJOTE simulations with non-Gaussian initial conditions [35]. Finally, it was shown that the local shape remains significantly affected even at higher redshifts, where the field is more linear but where we also expect to access more squeezed triangle configurations [36]. In order to improve constraints on pnGs from the power spectrum and bispectrum, we need to find ways to reduce the effects of mode coupling. This is the topic of the rest of this chapter.

4.3   Reconstructing the initial conditions

Since the coupling of modes is a consequence of the non-linear gravitational evolution of the density field, we expect to be able to access more information by reversing the gravitational collapse. This idea of *reconstruction* has been studied extensively in the context of BAOs, where it is used to recover the BAO peak in the two-point correlation function that is otherwise washed out due to the gravitational collapse [38]. It has been shown that this BAO reconstruction method can aid in reducing non-Gaussian covariance and improving parameter constraints [39], especially in the context of primordial non-Gaussianity [40]. The latter work only considered the tree-level primordial contribution (4.5) to the bispectrum, whereas in the non-linear regime, we expect additional loop contributions to both the power spectrum and bispectrum. Since reconstruction affects the loop contributions (and thereby the information contained in the power spectrum and bispectrum), any assessment of the improvement due to reconstruction should include the full non-linear power spectrum and bispectrum. Furthermore, we expect to be able to improve constraints using more sophisticated reconstruction methods than the standard reconstruction algorithm.

4.3.1   Reconstruction Algorithms

Although a full review of existing reconstruction algorithms is beyond the scope of this work (see e.g. [57]), we briefly summarize here several methods and their performance for completeness and comparison. Standard reconstruction (or Lagrangian-Growth-Shift), as originally developed for reconstructing the BAO feature in the matter power spectrum, relies on approximating the linear displacement field in order to move structure back to their initial positions [38]. When performed on a redshift $z = 0$ N-body simulation snapshot, the reconstructed density field correlates up to $\sim 37\%$ with the initial conditions at $k = 0.4\,\mathrm{h/Mpc}$ (see Figure 4.3). Since this original proposal, more sophisticated methods have been developed both at the level of the object (e.g. halo) catalog (Lagrangian space) [46] as well as the density field directly (Eulerian space) [57]. The iterative reconstruction algorithm of Ref. [46] can be considered the state-of-the-art non-machine learning approach, with a reconstruction cross-correlation of $\sim 90\%$ out to scales $k \leq 0.4\,\mathrm{h/Mpc}$ at redshift $z = 0$.

4.3.2   Reconstruction using U-Nets

In this work, we will focus on the application of neural networks in the context of reconstruction. Previous work has shown that a combination of standard reconstruction and convolutional neural networks is able to reconstruct the initial density field

significantly better than standard reconstruction alone [45]. Instead, we defer the entire reconstruction to a neural network applied directly to the density field. In particular, we will employ a U-Net, which was originally developed for medical image segmentation [58]. Recently, there have been many successful applications of U-Nets in cosmology and astrophysics addressing problems that involve the coupling of different length scales (see e.g. [59, 60]). Ordinary convolutional neural networks have a relatively small receptive field (depending on the size of the filter and the number of convolutions), making them local models. U-Nets incorporate down-sampling, effectively resulting in a larger receptive field, allowing them to also access non-local (large-scale) information, which is essential for our application.

Network architecture

The architecture used in this work resembles the U-Net developed in Ref. [60]. The network takes a density field on a 256^3 grid. Inspired by grid-based perturbation theory [61], we compute the following velocity fields on-the-fly:

$$\boldsymbol{u}(\mathbf{x}) = \int d^3\mathbf{k} \, \frac{-i\mathbf{k}}{k^2} \delta_m(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{4.6}$$

$$\partial_i(\boldsymbol{u})_j(\mathbf{x}) = \int d^3\mathbf{k} \, \frac{k_i k_j}{k^2} \delta_m(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{4.7}$$

yielding six additional input fields. Subsequently, we apply an ordinary convolutional layer with a kernel of size 3^3, followed by a *context block*, consisting of two additional convolutional layers of size 3^3, after which we add the output of the initial convolution to the output of the context block, this is known as a *residual connection*. We repeat this, only now employing a stride of two in the initial convolutional layer, effectively down-sampling the feature maps by a factor of two in every spatial dimension. This is then repeated three more times until the feature maps have a size of 16^3 cells. We then upsample the feature maps by a factor of two and concatenate the output of the residual connections in the down-sampling part of the network, forming what is known as *skip connections*. The concatenated feature maps are fed into a *localization block*, consisting of an ordinary convolutional layer of size 3^3 followed by a convolutional layer of size 1^3. This is repeated until we have reached the input size of 256^3, after which we perform one more convolution of size 1^3, in order to output a single feature map, representing the reconstructed density field. All but the last convolutional layer consists of the convolution, *instance normalization*, and a *Leaky ReLU* activation. The complete architecture is visualized in Figure 4.1. The network has roughly 2 million parameters implemented using Keras/Tensorflow and is publicly

available including trained weights.[1]
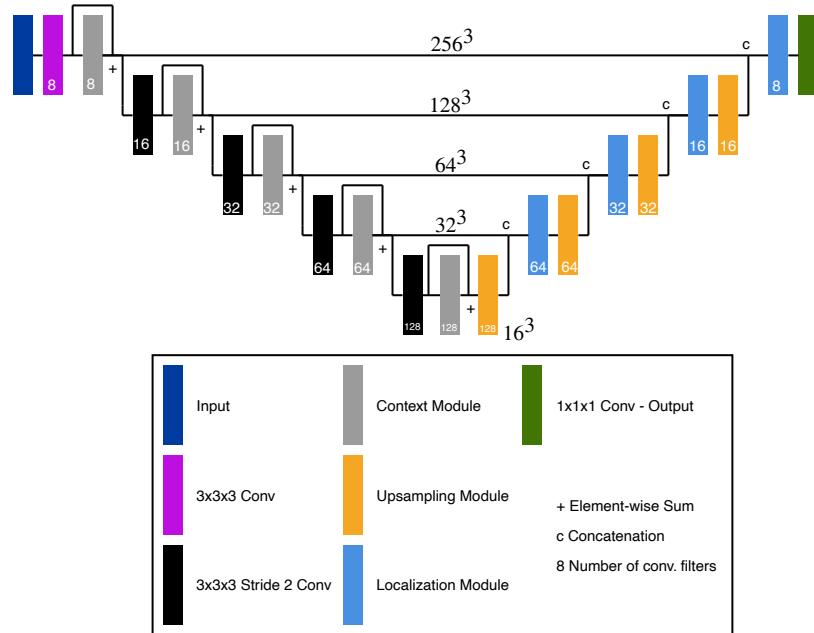
Figure 4.1: *Diagrammatic representation of the neural network architecture used in this chapter. The network is based on the one presented in Ref. [60].*

Training

The training dataset consists of 48 pairs of initial ($z = 127$) and final ($z = 0$) density fields of the fiducial cosmology QUIJOTE simulations, generated from the dark matter particle snapshots using a Piecewise Cubic Spline (PCS) mass assignment scheme to a grid of 256^3 cells.[2] The aim of our network is to minimize the loss function that is the mean squared error (MSE) between reconstructed and initial density fields:

$$L_{\text{MSE}} = \frac{1}{N_{\text{sims}}} \sum_{i}^{N_{\text{sims}}} \frac{1}{N_{\text{cells}}} \sum_{\mathbf{x}} \left(\delta_{\text{recon}}^{(i)}(\mathbf{x}) - \delta_{\text{init}}^{(i)}(\mathbf{x}) \right)^2, \tag{4.8}$$

---

[1] https://github.com/tsfloss/URecon

[2] This might seem like a small dataset, but every simulation constitutes 256^3 data points that need to be mapped to equally as many outputs. Training can be done with an even smaller dataset than the one considered here.

where $\delta_{\mathrm{recon}}^{(i)}$ and $\delta_{\mathrm{init}}^{(i)}$ are the reconstructed and initial ($z = 127$) density field of simulation i. It can be shown that the Fourier transform of this loss function is the mean squared error between different wavelength modes of the density field [45].

We will eventually be interested in determining the information content of cosmological parameters of the power spectrum and bispectrum of the reconstructed density field. The FFT-based estimator used to compute the bispectrum can only be used up to a scale of $k_{\mathrm{max}} = \frac{2}{3}k_{\mathrm{Nyquist}} \approx 0.53$ h/Mpc [62]. We have also explicitly checked that the power spectrum and bispectrum measurements up to this scale do not suffer from any aliasing effects, by comparing them to measurements made at a higher resolution grid of 1024^3 cells. Since we are only interested in a sub-sample of the modes present in our density field, we limit our target initial density fields already to this same k_{max} using an isotropic cutoff in Fourier space. This has the advantage that the network does not have to reconstruct modes that we will not use in our final analysis, speeding up the training process. We train the network using 4 Nvidia A100 40GB GPUs until we see no more significant improvement in the validation loss, taking a bit under 4 hours.

Network performance

We assess the quality of the reconstruction by computing the cross-correlation between two fields:

$$C_{X,Y}(\mathbf{k}) = \frac{\langle \delta_{\mathrm{X}}(\mathbf{k})\delta_{\mathrm{Y}}^*(\mathbf{k})\rangle}{\sqrt{P_{\mathrm{X}}(\mathbf{k})P_{\mathrm{Y}}(\mathbf{k})}}, \tag{4.9}$$

where $P_X(\mathbf{k})$ is the power spectrum of the field X and X, Y can be either the reconstructed, initial ($z = 127$) or final ($z = 0$) density fields. A perfect reconstruction then corresponds to $C_{\mathrm{recon,init}}(\mathbf{k}) = 1$. In Figure 4.2 we show an example of the network input and output and its corresponding cross-correlations are shown in Figure 4.3a, demonstrating the ability of our network to reconstruct the density field with a cross-correlation of $\sim 75\%$ up to the smallest scales of interest in this analysis. We also show the results if we use lower resolution density fields of 128^3 cells, demonstrating that the main limitation to further improve the reconstruction is the resolution of the density fields, as the reconstruction benefits from having access to more small-scale modes that are coupled to larger ones that are to be reconstructed. We are currently limited to 256^3 density fields due to memory limitations on our GPUs.[3]

---

[3]The approach taken in Ref. [45] circumvents this issue by first performing standard reconstruction on a box of higher resolution and then reconstructing subboxes of smaller size. Since their analysis is performed at $z = 0.5$, in Figure 4.3 we also plot our network's performance at that redshift for comparison.

Furthermore, we show the results for a similar network that does not use the velocity field information, demonstrating that including velocity is advantageous. Finally, we show the performance of our network when the $z = 0$ density field contains an increased amount of shot-noise (we have generated the same density field using an 8 times smaller subsample of the dark matter particles), which degrades the reconstruction as expected [63]. Compared to the state-of-the-art iterative reconstruction algorithm of Ref. [46], our network achieves a similar accuracy (cross-correlation) of $\sim 90\%$ up to scales $k \leq 0.4$ h/Mpc, while only having access to modes up to the Nyquist frequency of the 256^3 grid, $k_{\mathrm{Nyq}} \approx 0.8$ h/Mpc. To further assess the ability of our network to 'linearize' the density fields, we can consider the output power spectra. In Figure 4.3b we show the mean power spectra over 12500 simulation boxes before and after reconstruction. We see that the reconstructed power spectrum more closely follows the linear power spectrum, especially recovering the BAO peaks. It is worth emphasizing that once trained, our network is able to perform reconstruction in under a second of time per field. Furthermore, our method works directly at the level of the density field, without having to perform computations on large object catalogs. This allows us to quickly reconstruct the large dataset necessary for our subsequent analysis of the information content of the pre and post-reconstructed power spectrum and bispectrum.

4.4 Information content of the reconstructed field

In order to determine the merit of our reconstruction for cosmological parameter inference, we perform a Fisher analysis of the information content of the power spectrum and bispectrum using the QUIJOTE simulation suite. These simulations allow us to work with the full non-linear power spectrum and bispectrum within the non-linear regime, giving us a reliable estimate of the improvements of parameter constraints, without the need for complicated models such as perturbation theory or effective field theory.

4.4.1 Power spectrum and bispectrum estimators

We measure the binned power spectrum from density fields $\delta(\mathbf{x})$ according to:

$$\hat{P}(k_i) = \frac{1}{N_i} \sum_{\mathbf{q} \in k_i} \delta(\mathbf{q})\delta^*(\mathbf{q}), \tag{4.10}$$

Note that these results have been obtained using different N-body simulations, fiducial cosmology, and resolution and should therefore only be compared to ours qualitatively.
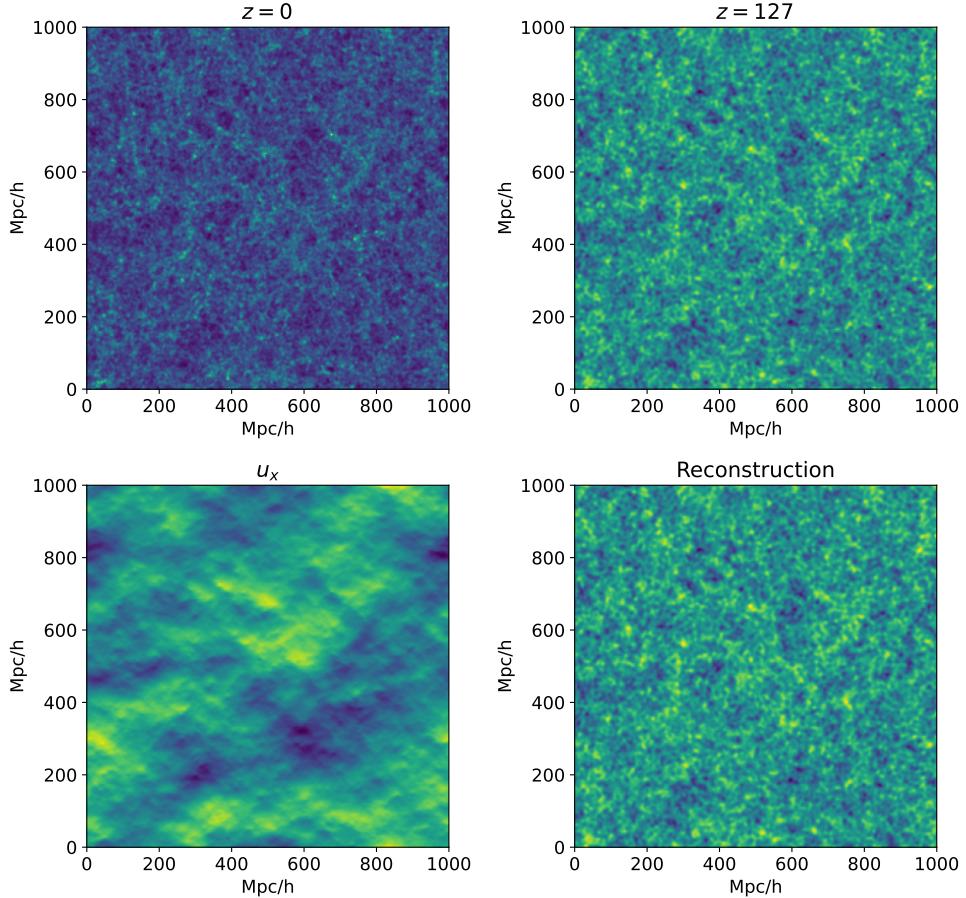
Figure 4.2: *Top: a validation pair of final ($z = 0$) and initial ($z = 127$) density field. Bottom left: one of the velocity fields generated on the fly. Bottom right: the corresponding reconstruction produced by our network. These are averaged along one entire spatial dimension (1000 Mpc/h) for illustrative purposes, whereas the actual fields are three-dimensional.*

where the sum runs over all momenta $\mathbf{q}$ that are within the shell $[k_i - \Delta k/2, k_i + \Delta k/2)$ and N_i is a normalization factor that counts the number of modes that fall in the bin. Similarly, for the bispectrum we compute

$$\hat{B}(k_1, k_2, k_3) = \frac{1}{N_{123}} \sum_{\mathbf{q}_1 \in k_1} \sum_{\mathbf{q}_2 \in k_2} \sum_{\mathbf{q}_3 \in k_3} (2\pi)^3 \delta_D^{(3)}(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3) \delta(\mathbf{q}_1) \delta(\mathbf{q}_2) \delta(\mathbf{q}_3),$$

(4.11)
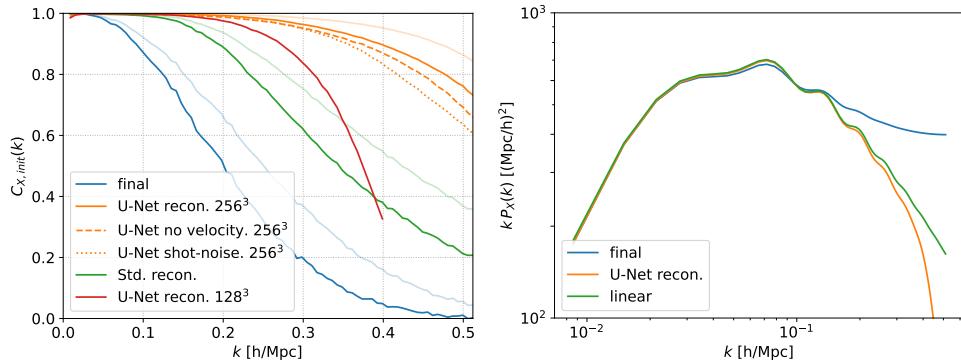
Figure 4.3: *Left: Cross correlations of various density fields. Blue shows the cross-correlation of the final ($z = 0$) and initial ($z = 127$) density field. Solid orange shows the cross-correlation of the reconstructed and initial density field. Green shows the cross-correlation of the density field reconstructed using standard reconstruction with a smoothing scale of $\sigma = 10\ h/Mpc$. Red shows the cross-correlation that can be achieved with a similar network and density fields of only 128^3 cells. Dashed orange shows the performance of a network that only takes the density field as an input, without velocity fields. Dotten orange shows the performance of our network when the final density field contains increased shot noise (generated with 8 times fewer dark matter particles). Transparent lines show the same results but for final density fields at $z = 0.5$ and can therefore be compared to Ref. [45]. Right: mean power spectrum over 12500 simulations before and after reconstruction, compared to the linear power spectrum (note that we multiplied by an additional factor of k to emphasize the BAO peaks).*

where N_{123} counts the number of triangle configurations within the bin. The bispectrum can be efficiently computed using fast Fourier transforms (FFTs). For the power spectrum, we use bins of width $\Delta k = k_F$ starting at the fundamental mode $k_{\min} = k_F$. The bispectrum is binned with width $\Delta k = 3k_F$, the first bin starting at $k_{\min} = 1.5k_F$. The smallest scale included in our analysis is $k_{\max} = 82.5k_F \approx 0.52\ \text{h/Mpc}$ resulting in 2276 triangle bins. We apply the usual compensation for the mass assignment (PCS) window function in Fourier space before measuring our statistics [64]. Measuring the power spectrum and bispectrum of the pre and post-reconstructed density fields, we obtain our data products $\boldsymbol{D} = \{P_{\text{pre}}, B_{\text{pre}}, P_{\text{post}}, B_{\text{post}}\}$.[4]

---

[4]We have used our own code that was cross-validated with other bispectrum codes. Our code is publicly available at `https://github.com/tsfloss/DensityFieldTools`

4.4.2 Covariance

An important quantity for determining the information content is the covariance matrix, which encodes correlations between data points. The covariance matrix for a data vector $\boldsymbol{D}_n$ (in our case the power spectrum and bispectrum measurements) measured from simulation n, is given by:

$$C_{ij} = \frac{1}{N-1} \sum_n^N \left(\boldsymbol{D}_n - \bar{\boldsymbol{D}}\right)_i \left(\boldsymbol{D}_n - \bar{\boldsymbol{D}}\right)_j, \tag{4.12}$$

where N is the number of measurements/simulations and $\bar{\boldsymbol{D}}$ is the mean of the data over all N simulations. We obtain the reconstructed covariance matrix by measuring the power spectrum and bispectrum of 12500 reconstructed QUIJOTE fiducial cosmology density fields at $z = 0$, which we demonstrate to be sufficient in Appendix 4.B. In order to visualize the correlation between different power spectrum and bispectrum bins, we use the correlation matrix:

$$r_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}. \tag{4.13}$$

In Figure 4.4 we show the correlation matrices of the pre-and post-reconstructed power spectrum and bispectrum. In this figure, the triangle bins have been ordered with increasing smallest momenta (largest scale), which emphasizes the fact that before reconstruction, triangles sharing the same shortest side are severely correlated, especially when very squeezed [34]. After reconstruction, modes have become significantly less correlated, indicating that additional information from higher-order correlation functions has been brought into the reconstructed power spectrum and bispectrum. Since non-Gaussian covariance is responsible for the saturation of parameter constraints in the non-linear regime [34–36], we expect that the reduction of covariance due to our reconstruction method will translate into improved parameter constraints.
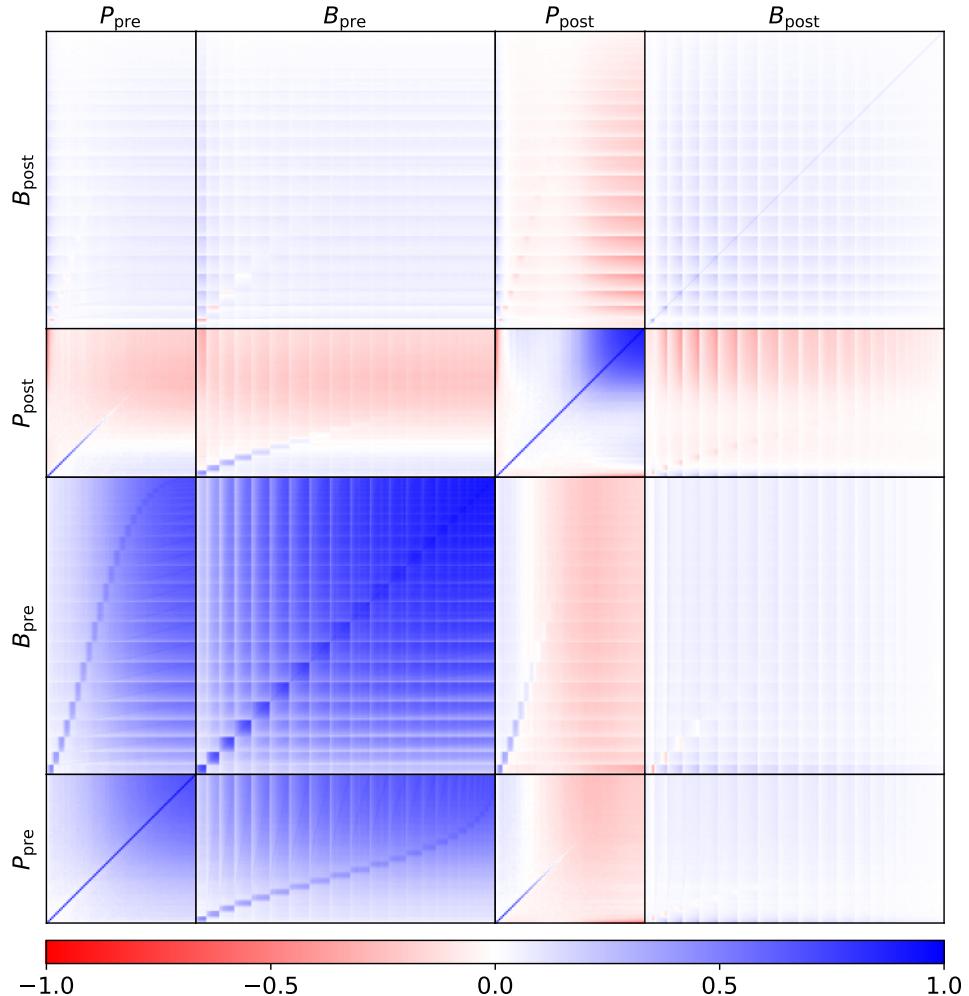
Figure 4.4: *Correlation matrix r_{ij} for the pre and post-reconstructed power spectrum and bispectrum. Triangle configurations for the bispectrum are ordered with increasing smallest momentum, visually creating blocks of triangles sharing the same shortest side in k-space.*

	$f_{\mathrm{NL}}^{\mathrm{loc}}$	$f_{\mathrm{NL}}^{\mathrm{equil}}$	$f_{\mathrm{NL}}^{\mathrm{orth}}$	h	n_s	Ω_m	Ω_b	σ_8
θ^{fid}	0	0	0	0.6711	0.9624	0.3175	0.049	0.834
$\delta\theta$	±100	±100	±100	±0.02	±0.02	±0.01	±0.002	±0.015

Table 4.1: *Cosmological parameters of the* QUIJOTE *simulations used in this chapter. The top row gives the fiducial cosmology used for training the neural network and computing the covariance matrix. The bottom row gives the variation of the parameters in simulations used to compute the derivatives in equation* (4.15)

4.4.3 Fisher Analysis

We determine the constraining power of the reconstructed power spectrum and bispectrum by performing a Fisher analysis. Under the assumption of a Gaussian likelihood, which we expect to be accurate enough for our purpose [65], the Fisher information matrix for parameters θ_a and data $\bar{D}$ (e.g. power spectrum and/or bispectrum) is given by:

$$F_{ab} = \frac{\partial \bar{D}}{\partial \theta_a} \cdot \left(C^{-1} \right) \cdot \frac{\partial \bar{D}}{\partial \theta_b}, \tag{4.14}$$

where the dot product runs over the data points. The inverse covariance matrix or precision matrix C^{-1}, is obtained by numerically inverting the covariance matrix of the previous section. The inversion of covariance matrices is biased by the limited amount of realizations available to construct the covariance matrix. We unbias our precision matrix by including the Hartlap factor [66]. In order to obtain the derivatives with respect to the cosmological parameters that we would like to constrain, we make use of the QUIJOTE simulations that vary $\{f_{\mathrm{NL}}^{\mathrm{local}}, f_{\mathrm{NL}}^{\mathrm{equil}}, f_{\mathrm{NL}}^{\mathrm{orth}}, h, n_s, \Omega_m, \Omega_b, \sigma_8\}$ by a fixed amount around their fiducial values, i.e. $\theta_a^{\mathrm{fid}} \pm \delta\theta_a$ as summarized in Table 4.1 [35, 47]. We reconstruct all these simulations (500 per varied parameter) and measure their power spectra and bispectra. We then compute the parameter derivatives by central differencing:

$$\frac{\partial \bar{D}}{\partial \theta_a} = \frac{\bar{D}_{\theta_a^{\mathrm{fid}}+\delta\theta_a} - \bar{D}_{\theta_a^{\mathrm{fid}}-\delta\theta_a}}{2\delta\theta_a}. \tag{4.15}$$

Here $\bar{D}_{\theta_a^{\mathrm{fid}}\pm\delta\theta_a}$ presents the mean measured data from the simulations run with cosmological parameter $\theta_a = \theta_a^{\mathrm{fid}} \pm \delta\theta_a$. The Fisher matrix allows us to compute the marginalized error on model parameters:

$$\sigma_a = \sqrt{(F^{-1})_{aa}}, \tag{4.16}$$

as well as their correlation coefficient, quantifying the amount of degeneracy between parameters:

$$\rho_{ab} = \frac{(F^{-1})_{ab}}{\sqrt{\sigma_a^2 \sigma_b^2}}. \tag{4.17}$$

Furthermore, the Fisher matrix can be used to construct an unbiased, minimal-variance marginalized estimator that estimates the parameter θ_a from the observed data $\boldsymbol{D}^{\text{obs}}$ given a fiducial model $\bar{\boldsymbol{D}}^{\text{fid}}$:

$$\hat{\theta}_a - \theta_a^{\text{fid}} = \sum_b (F^{-1})_{ab} \frac{\partial \bar{\boldsymbol{D}}}{\partial \theta_b} \cdot \left(C^{-1}\right) \cdot \left(\boldsymbol{D}^{\text{obs}} - \bar{\boldsymbol{D}}^{\text{fid}}\right). \tag{4.18}$$

The expected error of the estimate is equal to the Fisher error $\langle \hat{\theta}_a^2 \rangle = \sigma_a^2 = (F^{-1})_{aa}$. This estimator marginalizes over all parameters to make sure that parameter discrepancies between our fiducial model and the real cosmology do not translate into potentially large biases of the estimated parameter. Since our network is trained only on the fiducial cosmology simulations, reconstructions of simulations with different parameters are biased. Since we use these biased reconstructions to compute the derivatives in equation (4.15), this bias is thus included in the marginalization process. We therefore expect estimates of primordial non-Gaussianity to be unbiased and minimal-variance, even when the cosmology of the measured data does not match the fiducial cosmology.

4.4.4 Results

From the calculation of the Fisher information matrix of the pre and post-reconstructed power spectrum and bispectrum, Figure 4.5 shows the resulting confidence intervals and marginalized errors of all 8 parameters when fitting for all of them simultaneously. The results are based on using both power spectrum and bispectrum data (see Appendix 4.A for the results when using the power spectrum and bispectrum separately). The results before reconstruction match those presented in Ref. [35]. Most notably, after reconstruction there is a significant improvement in the error of all parameters when combining the pre- and post-reconstructed statistics as summarized in Table 4.2. To understand the improvement of these constraints, Figure 4.6 shows the correlation coefficient between the parameter constraints, revealing a drastic reduction in degeneracy model parameters. In Table 4.3, we present the (un-)marginalized constraints on f_{NL}. We conclude that our reconstruction method improves marginalized constraints on f_{NL} by a factor of 3.65, 3.54, and 2.90 for local, equilateral, and orthogonal respectively. The improved constraints are attributed to the reduced covariance and parameter degeneracy. Our results show that in order to

	$f_{\mathrm{NL}}^{\mathrm{local}}$	$f_{\mathrm{NL}}^{\mathrm{equil}}$	$f_{\mathrm{NL}}^{\mathrm{orth}}$
P_{pre}^{*}	2884 (31.31)	7888 (79.41)	3478 (187.2)
P_{post}	1890 (69.11)	3500 (165.5)	572.8 (116.4)
$P_{\mathrm{pre}} + P_{\mathrm{post}}$	135.6 (29.06)	481.2 (76.94)	175.4 (77.70)
B_{pre}	101.8 (42.65)	187.7 (123.8)	83.39 (61.39)
B_{post}	18.35 (16.86)	72.63 (34.25)	26.76 (16.85)
$B_{\mathrm{pre}} + B_{\mathrm{post}}$	16.94 (15.61)	60.65 (33.05)	22.03 (15.51)
$(P + B)_{\mathrm{pre}}$	51.64 (24.45)	159.3 (59.77)	49.08 (42.56)
$(P + B)_{\mathrm{post}}$	17.10 (15.61)	64.73 (32.58)	22.58 (15.49)
$(P + B)_{\mathrm{pre}} + (P + B)_{\mathrm{post}}$	14.14 (12.77)	45.01 (30.21)	16.93 (13.90)

Table 4.3: *Marginalized (Unmarginalized) errors on f_{NL} pre, post and pre+post reconstruction using different data products for a dark matter survey with volume $V = 1$ (Gpc/h)^3 at redshift $z = 0$ and $k_{\mathrm{max}} \approx 0.52$ h/Mpc. * Note that the parameter constraints from the pre-reconstructed power spectrum alone are likely too optimistic since the Fisher derivatives have not converged, as discussed in Appendix 4.B and Ref. [35].*

get an accurate estimate of the improvement on f_{NL} it is important to include both the power spectrum and bispectrum and to marginalize over other cosmological parameters. Naively, using only post-reconstructed bispectrum measurements (B_{post}) would lead to an improvement of marginalized constraints by a factor of 5.55 (local), 2.58 (equilateral) and 3.12 (orthogonal). Similarly, using only the tree-level primordial bispectrum contribution in equation (4.5), as was done in Ref. [40], suggests improvements on marginalized constraints by a factor 11.69 (local), 4.76 (equilateral) and 2.14 (orthogonal), demonstrating the necessity of including the full non-linear power spectrum and bispectrum in order to arrive at realistic estimates. Finally, estimating f_{NL} from the (reconstructed) QUIJOTE simulations using the estimator in equation (4.18) with the different products, we confirm that the estimator is unbiased and minimum-variance.

$f_{\mathrm{NL}}^{\mathrm{local}}$	$f_{\mathrm{NL}}^{\mathrm{equil}}$	$f_{\mathrm{NL}}^{\mathrm{orth}}$	h	n_s	Ω_m	Ω_b	σ_8
3.65	3.54	2.90	2.43	1.98	1.64	2.56	1.91

Table 4.2: *Improvement factor of parameter constraints when using the full data product $(P + B)_{\mathrm{pre}} + (P + B)_{\mathrm{post}}$ as compared to just $(P + B)_{\mathrm{pre}}$*
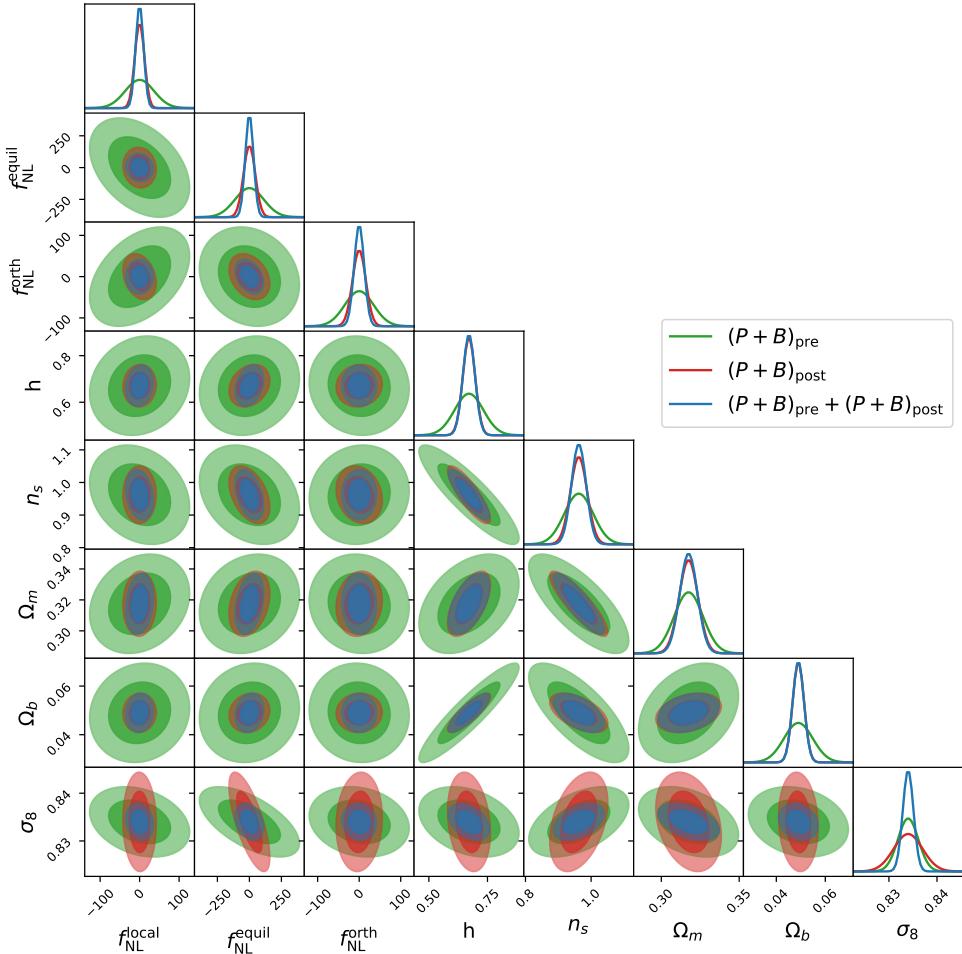
**Figure 4.5:** *1 and 2-σ marginalized constraints when jointly fitting for all eight parameters using the power spectrum and bispectrum* pre*,* post *and* pre+post *reconstruction.*

4.5   Conclusions

In this work, we have studied the use of neural networks in reconstructing the initial conditions from the late-time density field, with the main goal of improving future constraints on primordial non-Gaussianity. We have demonstrated that a U-Net architecture is able to reconstruct the initial conditions with an accuracy (cross-correlation) of $\sim 90\%$ up to scales $k \leq 0.4$ h/Mpc directly from the redshift $z = 0$ density field at a fraction of the computational time required by competitive meth-
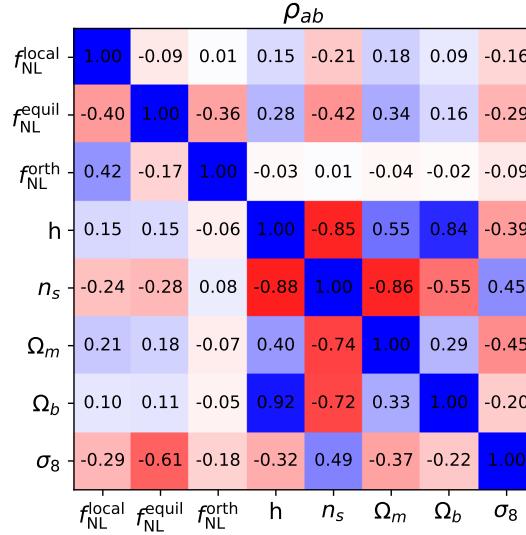
Figure 4.6: *Correlation coefficients $\rho_{ij} \in [-1, 1]$ between the different cosmological parameters when using the data products $(P + B)_{\mathrm{pre}}$ (lower triangular) and $(P + B)_{\mathrm{pre}} + (P + B)_{\mathrm{post}}$ (upper triangular).*

ods such as that of Ref. [46]. Furthermore, we have analyzed the information content of the dark matter density field after reconstruction by reconstructing a large part of the QUIJOTE simulation suite. From a Fisher analysis, we find that marginalized constraints on primordial non-Gaussianity from a joint analysis of power spectrum and bispectrum, are improved by a factor of 3.65, 3.54, 2.90 for local, equilateral, and orthogonal non-Gaussianity respectively. Given the strong saturation of these constraints in the non-linear regime due to non-Gaussian covariance [35], these constitute valuable improvements that cannot be realized by just including ~ 10 times more modes, as a naive mode-counting argument suggests. Hence, our work marks important progress towards a more optimal estimation of primordial non-Gaussianity.

It is important to note that we have not performed any extensive optimization of our network architecture or its hyperparameters, thus it is conceivable that the reconstruction performance can be further improved. Furthermore, our reconstruction is so far limited by the amount of memory available on GPUs.

This work can naturally be extended in a number of ways that we intend to explore in a future work:

1. In this work we have limited ourselves to the dark matter density field. In reality, we only have access to a biased tracer thereof, such as galaxies or atomic/molecular spectral lines (e.g. 21-cm). At low redshifts, these trace dark matter halos. To further assess the realism of this method we need to apply our methods to the dark matter halo field and eventually a real tracer.

2. In a real survey, we cannot observe the real position along the line-of-sight direction. Instead, we determine their position in redshift space, which is uncertain due to the unknown peculiar velocity of the objects. These *redshift space distortions* are an additional source of non-linearity. Again, instead, we should provide our network with late-time density fields in redshift space as input and train this to reconstruct the real space initial conditions.

Future constraints on primordial non-Gaussianity coming from large-scale structure will rely on our ability to model the non-linear nature of the tracer field. Traditionally, the emphasis in this field has been on modeling the statistical properties of the tracer field and trying to push these further into the non-linear regime in order to access more modes that can be compared to observation. Only more recently it has been shown that this approach shows diminishing returns as non-linear modes become increasingly correlated, reducing their information content. This is essentially a manifestation of the non-linear evolution scrambling the information of the linear field into higher-order correlation functions and additional work needs to be done to recover the information contained in the initial conditions. Therefore, it is important to emphasize that the complications due to mode coupling encountered in this work essentially turn N-point correlation functions into highly degenerate sub-optimal statistics. Although correlation functions are the natural object to study from a theoretical perspective, especially when it comes to primordial non-Gaussianity, there might be more optimal ways to compress survey data into statistics (see e.g. [67, 68]). It would be interesting to study alternative statistics in the context of reconstruction as well.

4.A   Information content of power spectrum and bispectrum separately

For completeness, we present in Figure 4.7 and Figure 4.8 the confidence intervals when using only the power spectrum or the bispectrum, respectively. By combining the pre and post-reconstructed power spectrum we gain significant information over either one of them alone, as was pointed out also in Ref. [39].
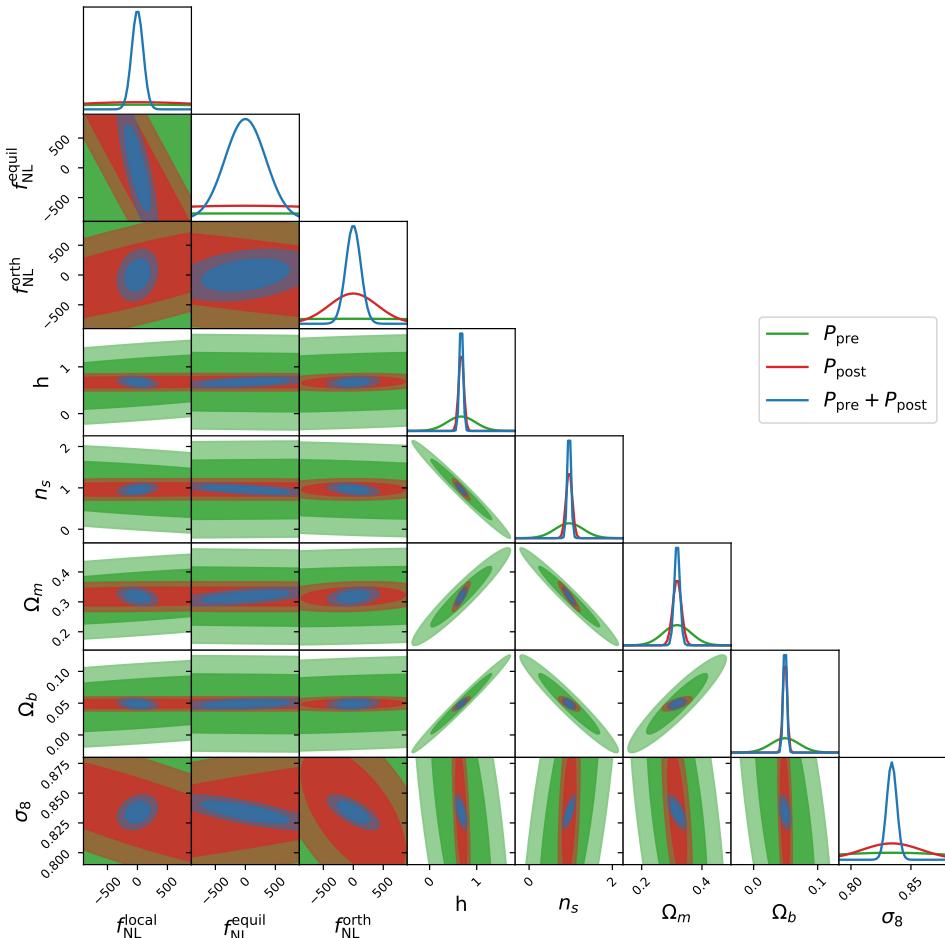

Figure 4.7: *1 and 2-σ marginalized constraints when jointly fitting for all eight parameters using the* pre*,* post *and* pre+post *reconstructed power spectrum. Note that the constraints from only the pre-reconstructed power spectrum are likely too optimistic since the Fisher derivatives have not fully converged, as discussed in Appendix 4.B and Ref. [35].*
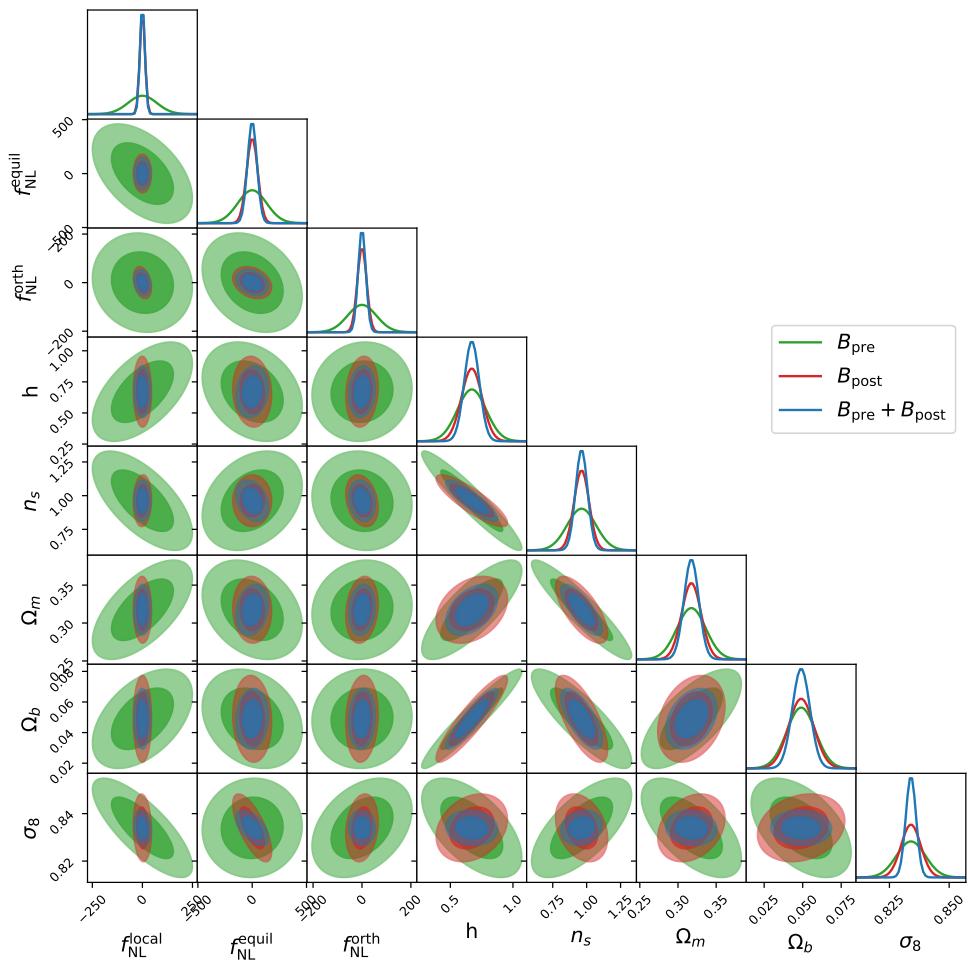
Figure 4.8: *1 and 2-σ marginalized constraints when jointly fitting for all eight parameters using the pre, post and pre+post reconstructed bispectrum.*

4.B   Convergence of Fisher analysis

We present here tests of the convergence of the Fisher forecasts presented in the main text. First, we assess whether the covariance matrix has converged sufficiently to provide robust parameter constraints. In Figure 4.9, we plot the variation of the marginalized parameter constraints from a Fisher analysis, as a function of the number of simulations included in the computation of the covariance matrix, while keeping the number of simulations used to compute derivatives fixed at maximum (i.e. 500). As we can see, the parameter constraints from all data products are increasingly stable as we include more simulations and appear to have converged well at 12500 simulations. We thus conclude that we have used a sufficient number of simulations to compute the covariance matrix in our analysis.
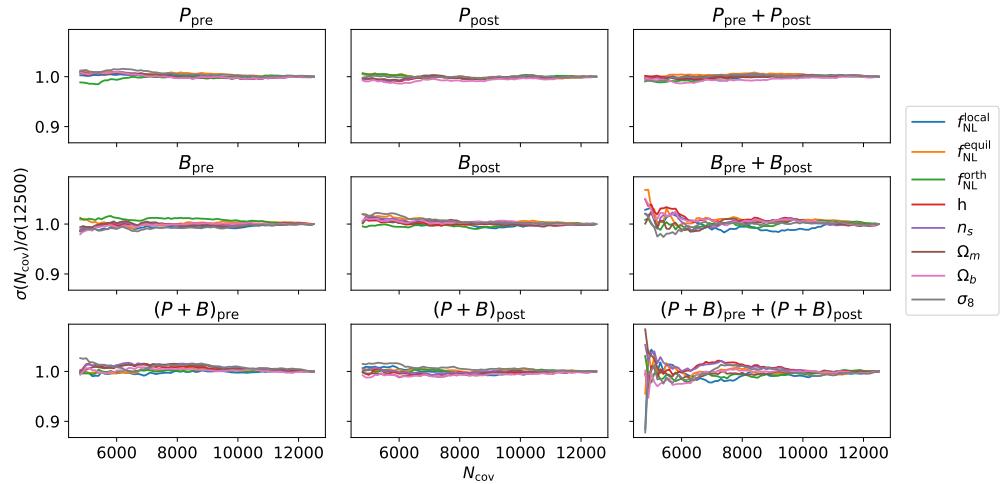

Figure 4.9: *Variation of the marginalized parameter constraints as a function of the number of simulations included in the computation of the covariance matrix, for all the different data products. For this test, we keep the number of simulations used to compute derivatives fixed at 500.*

Next, we test whether the derivatives that enter the Fisher matrix computation of equation (4.14), have converged. In Figure 4.10, we show the variation of the marginalized parameters constraints, as a function of the number of simulations included in the computation of the derivatives, while keeping the number of simulations used to compute the covariance matrix fixed at maximum (i.e. 12500). As was found in the original work of Ref. [35], the pre-reconstructed power spectrum has not converged and hence the forecasts based on this data alone should not be

trusted. However, the parameter constraints for all the other data products seem to have converged well when using 500 simulations to compute the derivatives.
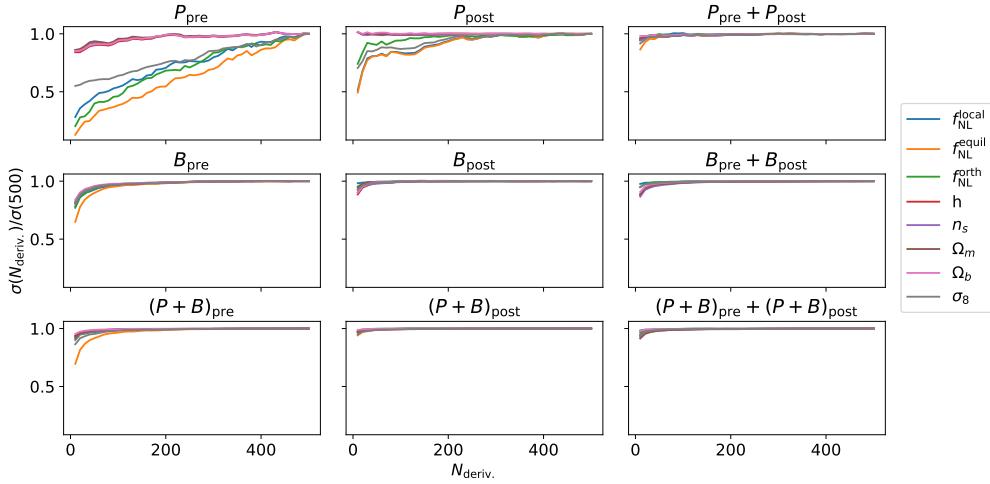

Figure 4.10: *Variation of the marginalized parameter constraints as a function of the number of simulations included in the computation of the Fisher derivatives, for all the different data products. For this test, we keep the number of simulations used to compute the covariance matrix fixed at 12500.*

Bibliography

[1] A. Achúcarro *et al.*, "Inflation: Theory and Observations," 3 2022.

[2] P. D. Meerburg *et al.*, "Primordial Non-Gaussianity," 3 2019.

[3] N. Arkani-Hamed and J. Maldacena, "Cosmological Collider Physics," 3 2015.

[4] H. Lee, D. Baumann, and G. L. Pimentel, "Non-Gaussianity as a Particle Detector," *JHEP* **12**, p. 040, 2016.

[5] P. G. Ferreira, J. Magueijo, and K. M. Gorski, "Evidence for nonGaussianity in the DMR four year sky maps," *Astrophys. J. Lett.* **503**, pp. L1–L4, 1998.

[6] E. Komatsu, B. D. Wandelt, D. N. Spergel, A. J. Banday, and K. M. Gorski, "Measurement of the cosmic microwave background bispectrum on the COBE DMR sky maps," *Astrophys. J.* **566**, pp. 19–29, 2002.

[7] E. Komatsu *et al.*, "First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: tests of gaussianity," *Astrophys. J. Suppl.* **148**, pp. 119–134, 2003.

[8] P. Creminelli, A. Nicolis, L. Senatore, M. Tegmark, and M. Zaldarriaga, "Limits on non-gaussianities from wmap data," *JCAP* **05**, p. 004, 2006.

[9] P. A. R. Ade *et al.*, "Planck 2015 results. XVII. Constraints on primordial non-Gaussianity," *Astron. Astrophys.* **594**, p. A17, 2016.

[10] Y. Akrami *et al.*, "Planck 2018 results. IX. Constraints on primordial non-Gaussianity," *Astron. Astrophys.* **641**, p. A9, 2020.

[11] P. Ade *et al.*, "The Simons Observatory: Science goals and forecasts," *JCAP* **02**, p. 056, 2019.

[12] K. N. Abazajian *et al.*, "CMB-S4 Science Book, First Edition," 10 2016.

[13] A. Kalaja, P. D. Meerburg, G. L. Pimentel, and W. R. Coulton, "Fundamental limits on constraining primordial non-Gaussianity," *JCAP* **04**, p. 050, 2021.

[14] J. B. Muñoz, Y. Ali-Haïmoud, and M. Kamionkowski, "Primordial non-gaussianity from the bispectrum of 21-cm fluctuations in the dark ages," *Phys. Rev. D* **92**(8), p. 083508, 2015.

[15] P. D. Meerburg, M. Münchmeyer, J. B. Muñoz, and X. Chen, "Prospects for Cosmological Collider Physics," *JCAP* **03**, p. 050, 2017.

[16] D. Karagiannis, J. Fonseca, R. Maartens, and S. Camera, "Probing primordial non-Gaussianity with the power spectrum and bispectrum of future 21 cm intensity maps," *Phys. Dark Univ.* **32**, p. 100821, 2021.

[17] T. Flöss, T. de Wild, P. D. Meerburg, and L. V. E. Koopmans, "The Dark Ages' 21-cm trispectrum," *JCAP* **06**(06), p. 020, 2022.

[18] G. Cabass, M. M. Ivanov, O. H. E. Philcox, M. Simonovic, and M. Zaldarriaga, "Constraining Single-Field Inflation with MegaMapper," 11 2022.

[19] J. R. Pritchard and A. Loeb, "21-cm cosmology," *Rept. Prog. Phys.* **75**, p. 086901, 2012.

[20] E. D. Kovetz *et al.*, "Line-Intensity Mapping: 2017 Status Report," 9 2017.

[21] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro, "Large scale structure of the universe and cosmological perturbation theory," *Phys. Rept.* **367**, pp. 1–248, 2002.

[22] A. Taruya, F. Bernardeau, T. Nishimichi, and S. Codis, "RegPT: Direct and fast calculation of regularized cosmological power spectrum at two-loop order," *Phys. Rev. D* **86**, p. 103528, 2012.

[23] J. J. M. Carrasco, M. P. Hertzberg, and L. Senatore, "The Effective Field Theory of Cosmological Large Scale Structures," *JHEP* **09**, p. 082, 2012.

[24] A. Slosar, C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan, "Constraints on local primordial non-Gaussianity from large scale structure," *JCAP* **08**, p. 031, 2008.

[25] A. J. Ross *et al.*, "The Clustering of Galaxies in SDSS-III DR9 Baryon Oscillation Spectroscopic Survey: Constraints on Primordial Non-Gaussianity," *Mon. Not. Roy. Astron. Soc.* **428**, pp. 1116–1127, 2013.

[26] B. Leistedt, H. V. Peiris, and N. Roth, "Constraints on Primordial Non-Gaussianity from 800 000 Photometric Quasars," *Phys. Rev. Lett.* **113**(22), p. 221301, 2014.

[27] S. Ho *et al.*, "Sloan Digital Sky Survey III photometric quasar clustering: probing the initial conditions of the Universe," *JCAP* **05**, p. 040, 2015.

[28] E. Castorina *et al.*, "Redshift-weighted constraints on primordial non-Gaussianity from the clustering of the eBOSS DR14 quasars in Fourier space," *JCAP* **09**, p. 010, 2019.

[29] E.-M. Mueller *et al.*, "The clustering of galaxies in the completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Primordial non-Gaussianity in Fourier Space," 6 2021.

[30] G. Cabass, M. M. Ivanov, O. H. E. Philcox, M. Simonović, and M. Zaldarriaga, "Constraints on Single-Field Inflation from the BOSS Galaxy Survey," *Phys. Rev. Lett.* **129**(2), p. 021301, 2022.

[31] G. D'Amico, M. Lewandowski, L. Senatore, and P. Zhang, "Limits on primordial non-Gaussianities from BOSS galaxy-clustering data," 1 2022.

[32] G. Cabass, M. M. Ivanov, O. H. E. Philcox, M. Simonović, and M. Zaldarriaga, "Constraints on multifield inflation from the BOSS galaxy survey," *Phys. Rev. D* **106**(4), p. 043506, 2022.

4

[33] K. C. Chan and L. Blot, "Assessment of the Information Content of the Power Spectrum and Bispectrum," *Phys. Rev. D* **96**(2), p. 023528, 2017.

[34] M. Biagetti, L. Castiblanco, J. Noreña, and E. Sefusatti, "The covariance of squeezed bispectrum configurations," *JCAP* **09**, p. 009, 2022.

[35] W. R. Coulton, F. Villaescusa-Navarro, D. Jamieson, M. Baldi, G. Jung, D. Karagiannis, M. Liguori, L. Verde, and B. D. Wandelt, "Quijote-PNG: Simulations of Primordial Non-Gaussianity and the Information Content of the Matter Field Power Spectrum and Bispectrum," *Astrophys. J.* **943**(1), p. 64, 2023.

[36] T. Flöss, M. Biagetti, and P. D. Meerburg, "Primordial non-Gaussianity and non-Gaussian covariance," *Phys. Rev. D* **107**(2), p. 023528, 2023.

[37] S. Goldstein, A. Esposito, O. H. E. Philcox, L. Hui, J. C. Hill, R. Scoccimarro, and M. H. Abitbol, "Squeezing fNL out of the matter bispectrum with consistency relations," *Phys. Rev. D* **106**(12), p. 123525, 2022.

[38] D. J. Eisenstein, H.-j. Seo, E. Sirko, and D. Spergel, "Improving Cosmological Distance Measurements by Reconstruction of the Baryon Acoustic Peak," *Astrophys. J.* **664**, pp. 675–679, 2007.

[39] Y. Wang *et al.*, "Extracting high-order cosmological information in galaxy surveys with power spectra," 2 2022.

[40] M. Shirasaki, N. S. Sugiyama, R. Takahashi, and F.-S. Kitaura, "Constraining primordial non-Gaussianity with postreconstructed galaxy bispectrum in redshift space," *Phys. Rev. D* **103**(2), p. 023506, 2021.

[41] S. He, Y. Li, Y. Feng, S. Ho, S. Ravanbakhsh, W. Chen, and B. Póczos, "Learning to Predict the Cosmological Structure Formation," *Proc. Nat. Acad. Sci.* **116**(28), pp. 13825–13832, 2019.

[42] D. Jamieson, Y. Li, R. A. de Oliveira, F. Villaescusa-Navarro, S. Ho, and D. N. Spergel, "Field Level Neural Network Emulator for Cosmological N-body Simulations," 6 2022.

[43] D. Jamieson, Y. Li, S. He, F. Villaescusa-Navarro, S. Ho, R. A. de Oliveira, and D. N. Spergel, "Simple lessons from complex learning: what a neural network model learns about cosmic structure formation," 6 2022.

[44] V. Jindal, D. Jamieson, A. Liang, A. Singh, and S. Ho, "Predicting the Initial Conditions of the Universe using Deep Learning," 3 2023.

[45] C. J. Shallue and D. J. Eisenstein, "Reconstructing cosmological initial conditions from late-time structure with convolutional neural networks," *Mon. Not. Roy. Astron. Soc.* **520**(4), pp. 6256–6267, 2023.

[46] M. Schmittfull, T. Baldauf, and M. Zaldarriaga, "Iterative initial condition reconstruction," *Phys. Rev. D* **96**(2), p. 023505, 2017.

[47] F. Villaescusa-Navarro *et al.*, "The Quijote simulations," *Astrophys. J. Suppl.* **250**(1), p. 2, 2020.

[48] M. Biagetti, J. Calles, L. Castiblanco, K. González, and J. Noreña, "A Model for the Squeezed Bispectrum in the Non-Linear Regime," 12 2022.

[49] D. Babich, P. Creminelli, and M. Zaldarriaga, "The Shape of non-Gaussianities," *JCAP* **08**, p. 009, 2004.

[50] J. M. Maldacena, "Non-Gaussian features of primordial fluctuations in single field inflationary models," *JHEP* **05**, p. 013, 2003.

[51] P. Creminelli and M. Zaldarriaga, "Single field consistency relation for the 3-point function," *JCAP* **10**, p. 006, 2004.

[52] C. Cheung, P. Creminelli, A. L. Fitzpatrick, J. Kaplan, and L. Senatore, "The Effective Field Theory of Inflation," *JHEP* **03**, p. 014, 2008.

[53] L. Senatore, K. M. Smith, and M. Zaldarriaga, "Non-Gaussianities in Single Field Inflation and their Optimal Limits from the WMAP 5-year Data," *JCAP* **01**, p. 028, 2010.

[54] J. J. M. Carrasco, S. Foreman, D. Green, and L. Senatore, "The Effective Field Theory of Large Scale Structures at Two Loops," *JCAP* **07**, p. 057, 2014.

[55] R. E. Angulo, S. Foreman, M. Schmittfull, and L. Senatore, "The One-Loop Matter Bispectrum in the Effective Field Theory of Large Scale Structures," *JCAP* **10**, p. 039, 2015.

[56] W. R. Coulton, P. D. Meerburg, D. G. Baker, S. Hotinli, A. J. Duivenvoorden, and A. van Engelen, "Minimizing gravitational lensing contributions to the primordial bispectrum covariance," *Phys. Rev. D* **101**(12), p. 123504, 2020.

[57] M. Schmittfull, Y. Feng, F. Beutler, B. Sherwin, and M. Y. Chu, "Eulerian BAO Reconstructions and N-Point Statistics," *Phys. Rev. D* **92**(12), p. 123522, 2015.

[58] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention– MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pp. 234–241, Springer, 2015.

[59] T. L. Makinen, L. Lancaster, F. Villaescusa-Navarro, P. Melchior, S. Ho, L. Perreault-Levasseur, and D. N. Spergel, "deep21: a deep learning method for 21 cm foreground removal," *JCAP* **04**, p. 081, 2021.

[60] S. Gagnon-Hartman, Y. Cui, A. Liu, and S. Ravanbakhsh, "Recovering the Wedge Modes Lost to 21-cm Foregrounds," *Mon. Not. Roy. Astron. Soc.* **504**, p. 4716, 2021.

[61] A. Taruya, T. Nishimichi, and D. Jeong, "Grid-based calculation for perturbation theory of large-scale structure," *Phys. Rev. D* **98**(10), p. 103532, 2018.

4

[62] E. Sefusatti, M. Crocce, R. Scoccimarro, and H. Couchman, "Accurate Estimators of Correlation Functions in Fourier Space," *Mon. Not. Roy. Astron. Soc.* **460**(4), pp. 3624–3636, 2016.

[63] X. Chen, F. Zhu, S. Gaines, and N. Padmanabhan, "Effective cosmic density field reconstruction with convolutional neural network," *Mon. Not. Roy. Astron. Soc.* **523**(4), pp. 6272–6281, 2023.

[64] Y. P. Jing, "Correcting for the alias effect when measuring the power spectrum using FFT," *Astrophys. J.* **620**, pp. 559–563, 2005.

[65] R. Scoccimarro, "The bispectrum: from theory to observations," *Astrophys. J.* **544**, p. 597, 2000.

[66] J. Hartlap, P. Simon, and P. Schneider, "Why your model parameter confidences might be too optimistic: Unbiased estimation of the inverse covariance matrix," *Astron. Astrophys.* **464**, p. 399, 2007.

[67] G. Valogiannis and C. Dvorkin, "Towards an optimal estimation of cosmological parameters with the wavelet scattering transform," *Phys. Rev. D* **105**(10), p. 103534, 2022.

[68] G. Jung, D. Karagiannis, M. Liguori, M. Baldi, W. R. Coulton, D. Jamieson, L. Verde, F. Villaescusa-Navarro, and B. D. Wandelt, "Quijote-PNG: Quasi-maximum Likelihood Estimation of Primordial Non-Gaussianity in the Nonlinear Dark Matter Density Field," *Astrophys. J.* **940**(1), p. 71, 2022.

Chapter 5

Denoising Diffusion Delensing Delight

Abstract

Optimal extraction of cosmological information from observations of the Cosmic Microwave Background critically relies on our ability to accurately undo the distortions caused by weak gravitational lensing. In this work, we demonstrate the use of denoising diffusion models in performing Bayesian lensing reconstruction. We show that score-based generative models can produce accurate, uncorrelated samples from the CMB lensing convergence map posterior, given noisy CMB observations. To validate our approach, we compare the samples of our model to those obtained using established Hamiltonian Monte Carlo methods, which assume a Gaussian lensing potential. We then go beyond this assumption of Gaussianity, and train and validate our model on non-Gaussian lensing data, obtained by ray-tracing N-body simulations. We demonstrate that in this case, samples from our model have accurate non-Gaussian statistics beyond the power spectrum. The method provides an avenue towards more efficient and accurate lensing reconstruction, that does not rely on an approximate analytic description of the posterior probability. The reconstructed lensing maps can be used as an unbiased tracer of the matter distribution, and to improve delensing of the CMB, resulting in more precise cosmological parameter inference.

5.1 Introduction

The deflection of photons of the Cosmic Microwave Background (CMB) by the large-scale structure between today and the last-scattering surface, changes the CMB statistics, leaving a characteristic imprint [1–5]. This effect, known as weak gravitational lensing, affects our ability to infer properties of the early universe such as primordial gravitational waves through B-mode polarization [6], or the presence of primordial non-Gaussianity [7–9], both of which would constitute strong evidence for the inflationary paradigm and are thus important targets for experimental endeavors [10]. Furthermore, the gravitational potential that lenses the CMB is an unbiased tracer (albeit projected) of the total matter distribution of the universe and thus contains important cosmological information by itself. Accurately reconstructing the CMB lensing potential is therefore of considerable interest [11–14], for example, to constrain the mass of neutrinos [15], the late-time growth of structure (i.e σ_8) and local primordial non-Gaussianity, especially in cross-correlation with other tracers, through cosmic-variance cancellation [16, 17].

Several methods have been developed to reconstruct the lensing potential from observed CMB data, starting with the quadratic estimator by Seljak & Zaldarriaga [18], and Hu & Okamoto [19, 20]. These estimators have been used in virtually all CMB lensing analyses to date [21–26], for which they proved to be sufficiently optimal. The high resolution and low noise levels of current (South Pole Telescope 3G [27, 28]) and next generation (Simons Observatory [29] and CMB-S4 [30]) CMB surveys, allow sub-percent level lensing reconstruction and an unprecedented sensitivity to primordial B modes. At this point, however, the quadratic estimator ceases to be nearly optimal, and more sophisticated techniques are required to achieve optimal lensing reconstruction, such as iterative maximum a posteriori (MAP) estimators [31–34] and gradient inversion techniques [35]. Besides these analytical approaches, the use of machine learning methods in CMB lensing problems has been an active topic of research in recent years as well [36–39]. Although the quality of the reconstructed lensing potential with iterative or machine-learned estimators may be statistically optimal, both methods yield point estimates that require expensive simulations to propagate uncertainty in the estimates when they are used in any subsequent analyses (e.g. to infer primordial non-Gaussianity from delensed CMB data). Additionally, non-Gaussianity of the lensing potential has been shown to contain significant amounts of information on e.g. neutrino mass and dark energy [40]. Reconstructing the lensing convergence bispectrum is therefore of considerable interest. However, determining non-Gaussian statistics of the lensing potential using conventional estimators can be challenging, due to having to model complicated noise biases [41], although this may be somewhat alleviated with alternative estimators

[42].

More recently the problem of CMB delensing and lensing reconstruction has received a fully Bayesian treatment by Millea, Anderes & Wandelt [34, 43], which we will refer to as MAW (see also [44] for earlier work using only temperature modes). Their approach to this high-dimensional statistical problem relies on having access to the gradient of the CMB lensing posterior, allowing for the use of a Hamiltonian-Monte-Carlo-within-Gibbs sampler that simultaneously samples cosmological parameters (the amplitude A_ϕ of the lensing potential power spectrum, and the tensor-to-scalar ratio r), the delensed CMB data and the lensing potential. The ability to sample from the posterior enables the accurate determination of uncertainties when working with the delensed data or the inferred lensing potential. This powerful method has since been used to improve the analysis of data from SPT-3G [45].

One of the crucial assumptions for MAW being able to write down and evaluate the posterior is that the lensing potential is Gaussian. However, in reality, this assumption does not hold since the large-scale structure of the universe is non-Gaussian. More accurate reconstruction of the lensing potential will thus require going beyond this Gaussian *prior* assumption, but this is not straightforward with existing methods. At the same time, the Hamiltonian Monte Carlo (HMC) method typically requires long sampling chains to ensure convergence of the lensing maps, which scales inefficiently with the size of the data. Although this issue can be partially ameliorated with suitable approximations [46, 47], it is worth investigating alternative *simulation-based inference* solutions [48, 49].

Here, we explore the use of probabilistic machine learning models in reconstructing the lensing potential, alleviating some of the limitations of the HMC approach. Specifically, we consider the use of score-based generative models (SGM) [50], a variant of what is commonly referred to as *diffusion models*. Using techniques inspired by the stochastic nature of diffusion, these models can draw uncorrelated samples from complicated high-dimensional probability distributions, without requiring an explicit expression for this distribution, making them particularly suitable for problems involving images [51]. Simply put, such models learn to map samples from a simple distribution (i.e. a Gaussian) into samples of a more complicated distribution, iteratively turning pure noise into meaningful data. Diffusion models have been successfully applied to various problems in the context of astrophysics and cosmology (e.g. [52**? ?** –57]). Here we will demonstrate the use of such models in reconstructing the CMB lensing convergence map given noisy CMB observations, enabling sampling of the Bayesian lensing posterior. While MAW solves the problem with explicitly specified likelihood and priors, our method implements an implicit

inference approach where likelihood and prior are represented in terms of pairs of lensing and data maps. In our case these are generated through simulations, making our work an example of simulation-based inference. Because of this, we are able to reconstruct non-Gaussian lensing potentials with accurate non-Gaussian statistics for the first time, requiring only a change in training data. Our model, trained on non-Gaussian lensing maps from N-body simulations, generates samples with the correct non-Gaussian statistics, as demonstrated by the one-point PDF and bispectrum. On signal-dominated scales, the sampled maps show excellent agreement with the true non-Gaussian lensing potential, while on noise-dominated scales, they revert to the learned non-Gaussian prior. Notably, models trained only on Gaussian lensing maps fail to capture the full non-Gaussian structure. Furthermore, our method allows for fast and uncorrelated sampling of the lensing potential posterior, circumventing the convergence issues and inefficiencies faced by traditional Markov Chain Monte Carlo methods. These advancements open up new possibilities for efficient delensing and precise constraints on large-scale structure from upcoming CMB surveys.

The paper is organized as follows. In section 5.2 we go over the necessary basics of CMB lensing. In section 5.3 we introduce the Bayesian approach to CMB delensing by MAW. In section 5.4 we introduce some basic aspects of score-based generative diffusion models. To validate the use of these models for Bayesian lensing reconstruction tasks, we perform a detailed comparison against the method and code of MAW in section 5.5. Finally, we demonstrate that our model can be used for the accurate reconstruction of non-Gaussian lensing potentials in section 5.6.

5.2 CMB lensing basics

The intervening matter distribution gravitationally bends the paths of CMB photons that travel through it, such that we observe deflected photons. This induces a remapping of the true CMB temperature (T) and polarization (Stokes Q,U) modes to the observed, lensed quantities ($\tilde{T}, \tilde{Q}, \tilde{U}$) (for simplicity neglecting a small phase correction [58]):

$$
\begin{aligned}
\tilde{T}(\hat{\mathbf{n}}) &= T(\hat{\mathbf{n}} + \boldsymbol{\alpha}(\hat{\mathbf{n}})) \\
(\tilde{Q} \pm i\tilde{U})(\hat{\mathbf{n}}) &= (Q \pm iU)(\hat{\mathbf{n}} + \boldsymbol{\alpha}(\hat{\mathbf{n}}))
\end{aligned}
\tag{5.1}
$$

where $\hat{\mathbf{n}}$ is the unit vector that denotes the position in the sky, and $\boldsymbol{\alpha}$ is the deflection angle that encodes the remapping. This deflection angle is given by the gradient of the lensing potential ϕ:

$$
\boldsymbol{\alpha} = \nabla\phi,
\tag{5.2}
$$

which itself is a two-dimensional projection of the three-dimensional gravitational potential ψ:

$$\phi(\hat{\mathbf{n}}) = -2 \int d\chi \, \frac{\chi_{\mathrm{CMB}} - \chi}{\chi_{\mathrm{CMB}} \chi} \psi(\hat{\mathbf{n}}\chi, \chi), \tag{5.3}$$

where χ denotes the comoving distance, and χ_{CMB} is the comoving distance to the last-scattering surface, where the CMB was emitted (at redshift $z \approx 1100$). In the flat sky approximation, these fields can also be written in Fourier space, parametrized by wavevector $\boldsymbol{\ell}$. Another useful quantity is the the CMB lensing convergence κ:

$$\kappa(\hat{\mathbf{n}}) = -\frac{1}{2} \nabla^2 \phi(\hat{\mathbf{n}}), \tag{5.4}$$

which is most easily obtained in Fourier space as:

$$\kappa_{\boldsymbol{\ell}} = \frac{|\boldsymbol{\ell}|^2}{2} \phi_{\boldsymbol{\ell}}. \tag{5.5}$$

Using a Taylor expansion of Eq. (5.1) one can determine the lowest-order effect of lensing on the CMB fluctuations. This can then be used to construct a quadratic estimator for the CMB lensing convergence [18–20]. Additionally, iterating this estimator yields the maximum-a-posteriori estimate of the lensing convergence [31–33].

5.3 Bayesian delensing

In Refs. [34, 43], MAW present a Bayesian approach to lensing reconstruction and delensing of the CMB fluctuations. The approach takes as a starting point the posterior probability distribution of the true CMB f (i.e. T, Q, U), the lensing potential ϕ and cosmological parameters θ (the amplitude of the lensing power spectrum A_ϕ and the tensor-to-scalar ratio r), given observed data d (i.e. $\tilde{T}, \tilde{Q}, \tilde{U}$): $\mathcal{P}(f, \phi, \theta | d)$. This data constitutes lensed, noised, and masked CMB data:

$$d = \mathbb{A} \mathbb{L}(\phi) f + n \tag{5.6}$$

where $\mathbb{L}(\phi)$ denotes the lensing operation, n is the noise, and $\mathbb{A}$ is a linear transformation that includes instrumental effects such as the beam and sky mask. Using Bayes' theorem, this posterior can be expressed in terms of the likelihood of the data, and a prior (following the concise notation of MAW):

$$\mathcal{P}(f, \phi, \theta | d) \propto \mathcal{P}(d | f, \phi, \theta) \mathcal{P}(f, \phi, \theta). \tag{5.7}$$

Under the assumption that the noise is a Gaussian random field with covariance $\mathbb{C}_n$, the likelihood is proportional to:

$$\mathcal{P}(d | f, \phi, \theta) \propto \exp\left(-\frac{(d - \mathbb{A} \mathbb{L}(\phi) f)^2}{2\mathbb{C}_n} \right), \tag{5.8}$$

and the priors on the f and ϕ fields are taken to be Gaussian:

$$\mathcal{P}(f, \phi, \theta) \propto \exp\left(-\frac{f^2}{2\mathbb{C}_f}\right) \exp\left(-\frac{\phi^2}{2\mathbb{C}_\phi}\right) \mathcal{P}(\theta). \tag{5.9}$$

Since we are after the pixels of the fields f and ϕ, this constitutes a high-dimensional posterior distribution that is intractable with conventional Metropolis-Hastings Monte Carlo methods. MAW overcome this challenge by numerically implementing the posterior in a fully differentiable fashion (see CMBLENSING.JL [1]), providing access to the derivatives of the posterior with respect to all parameters (i.e. field pixels and θ). This enables both MAP estimation through gradient-based optimization [34] and Hamiltonian Monte Carlo sampling [43] of the posterior. Crucially, even if one is only after the lensing potential, they find that it is easier to sample the full posterior $\mathcal{P}(f, \phi, \theta | d)$, rather than the marginalized posterior $\mathcal{P}(\phi, \theta | d)$.

Although the Gaussian prior on the lensing potential does not imply that the posterior samples are necessarily Gaussian, it is not the maximally informative prior for data that have been lensed with a non-Gaussian potential, and we expect samples in this case to have inaccurate non-Gaussian statistics. Including such a non-Gaussian prior in the approach of MAW would require forward modeling and evaluation of the posterior, which is expensive and hard to implement. Additionally, HMC methods can be inefficient, due to subsequent samples being correlated. In practice, the sample chains of the large-scale modes of the lensing potential have long autocorrelation lengths, resulting in a low yield of independent samples of these scales, especially when a mask is included [33, 43].

In the rest of this work, we will demonstrate the use of generative machine learning models in reconstructing the CMB lensing convergence to overcome both of these limitations of the HMC approach, albeit at the expense of losing provable long-run convergence to the true posterior.

5.4   Score-based generative models

For our probabilistic machine learning approach, we adopt score-based generative models [50]. These models can produce samples of a learned data distribution $p(\mathbf{x})$ through a reverse-diffusion process. Starting from data $\mathbf{x}$, we incrementally perturb the data using Gaussian noise with an increasing variance schedule $\sigma^2(t)$ over many timesteps $t \in [0, 1]$, until it is pure noise. This *diffusion* process is described by a

---

[1] https://github.com/marius311/CMBLensing.jl

stochastic differential equation (SDE) [50]:

$$dx = f(x, t)dt + g(t)dw, \tag{5.10}$$

where f is called the drift coefficient, g is the diffusion coefficient, and dw denotes a Wiener (noising) process. Remarkably, the diffusion process has a reverse SDE that turns pure Gaussian noise into a sample of $p(x)$ by reversing the noising process at each timestep; the *denoising* process [59]:

$$dx = \left[f(x, t) - g(t)^2 \nabla_x p_t(x) \right] dt + g(t)d\bar{w}. \tag{5.11}$$

This requires access to the *score* $\nabla_x p_t(x)$, where $p_t(x)$ is the probability distribution of x at time t. Furthermore, if we are after the conditional distribution $p(x|y)$ (i.e. the posterior of parameters x given data y), we instead need the conditional score $\nabla_x p_t(x|y)$. In our case x represents the lensing convergence map, while y represents observed lensed CMB data (e.g. Q and U maps). We use a neural network, more specifically a U-Net, to approximate this conditional score by training it to predict random noise added to an image x at a random timestep, given conditioning data $y(x)$ [60].

Typically in diffusion models, the noising process dw uses white noise

$$z_w \sim \mathcal{N}(0, I), \tag{5.12}$$

but we find that both training and sampling are faster and more accurate when using *reddened* Gaussian noise with a power spectrum similar to the target κ maps:

$$z_\kappa(z_w) = \Sigma_\kappa^{1/2} z_w, \tag{5.13}$$

where Σ_κ is the theory covariance matrix of the lensing convergence maps, and the multiplication is performed in Fourier space, where this covariance matrix is diagonal (i.e. the power spectrum $C_\ell^{\kappa\kappa}$). This type of noise allows the model to affect all scales of the data at every timestep, instead of different scales at different timesteps, resulting in higher-quality samples with fewer diffusion steps. Furthermore, we employ the Variance Exploding variant of the algorithm [50]:

$$f(x, t) = 0, \qquad g(t) = \sqrt{\frac{d\sigma^2(t)}{dt}}, \qquad \sigma(t) = \sigma_{\min} \left(\frac{\sigma_{\max}}{\sigma_{\min}} \right)^t, \tag{5.14}$$

where $\sigma_{\min}$ and $\sigma_{\max}$ set the lowest and highest noise scales of the noising process. Given this setup, we use the noise schedule $\sigma_{\min} = 0.01$ and $\sigma_{\max} = 100$. This provides sufficient noise to properly swamp the image, as well as sufficient low-noise steps to fine-tune the sample. Finally, to not underemphasize small scales in the

model's training objective, the score-matching loss to be optimized by the network is still defined at the level of the white noise:

$$L = \mathbb{E}_{t,\mathbf{x}_0,\mathbf{z}_w} ||\mathbf{z}_w - \mathbf{s}_\theta(\mathbf{x}_0 + \sigma(t)\mathbf{z}_\kappa(\mathbf{z}_w), t)||^2 \tag{5.15}$$

where $\mathbf{s}_\theta(\mathbf{x}_t, t)$ is the U-Net's noise prediction at time t, given the noisy image $\mathbf{x}_t$.

After training, we sample from the learned conditional distribution by discretizing the reverse SDE with the Euler-Maruyama method:

$$\mathbf{x}_{t-\Delta t} = \mathbf{x}_t - g(t)^2 \left(\mathbf{\Sigma}_\kappa^{1/2}\, \mathbf{s}_\theta(\mathbf{x}_t, t)\right) \Delta t + g(t)\sqrt{\Delta t}\, \mathbf{z}_\kappa, \tag{5.16}$$

and applying this equation iteratively starting from pure noise at time $t = 1$. We use 1000 timesteps to draw samples, but good results can be achieved with fewer steps.

More details on the implementation of the model and U-Net can be found in Appendix 5.A.

5.5   Gaussian lensing reconstruction

To validate our method, we first perform a benchmark comparison against the HMC algorithm, using CMBLENSING.JL. As noted, this algorithm can be used to simultaneously sample the true CMB, the lensing potential ϕ, and parameters A_ϕ (the amplitude of the lensing potential power spectrum) and r (the tensor-to-scalar ratio). In this work, we will only focus on reconstructing the lensing potential, thus implicitly marginalizing over the true CMB. Ideally, we would marginalize over both the amplitude and tensor-to-scalar ratio. However, we found that the score-based generative algorithm, as well as variants thereof (e.g. Denoising Diffusion Probabilistic Models), struggle to yield samples with the correct amplitude if we vary the lensing amplitude. We comment more on this in Appendix 5.B. For now, we limit our setup to fixed amplitude $A_\phi = 1$, but we do marginalize over r.

We mimic one of the survey configurations in MAW's work and use CMBLENSING.JL to generate training data, consisting of 32768 pairs of Gaussian κ and lensed (Q, U)-maps, including masks and noise with specifications given in Table 5.1 and random $r \in [10^{-6}, 10^{-1}]$ (quadratically spaced). The (Q, U, κ)-maps are each normalized by the mean and standard deviation over the entire dataset. Furthermore, we apply a cutoff $\ell < 3000$ to the κ maps during training, so the network only learns to reconstruct modes below this cutoff (the (Q, U) maps still have $\ell < 5000$). In our setup, modes beyond this cutoff are severely noise-dominated in the reconstruction

Parameter	Configuration
CMB data	Q,U
Map size (pixels)	256×256
Pixel width	2 arcmin
Total area	73 deg^2
White noise level in P	1 μK-arcmin
$(\ell_{\text{knee}}, \alpha_{\text{knee}})$	$(100, 3)$
Beam FWHM	2 arcmin
Fourier masking	$2 < \ell < 5000$
Pixel masking	$0.4°$ border $+0.6°$ apod

Table 5.1: *Configuration used in our experiments. The setup follows that of the 2PARAM setup of [43].*

samples and therefore not informative. Finally, since these maps have been generated with periodic boundary conditions, we apply periodic padding in our U-Net.

We validate our generative model by comparing 2048 samples of our model and a CMBLENSING.JL chain of 17000 samples (we drop the first 1000 samples and keep every fifth sample, leaving us with 3200 samples from the chain), for a validation case with parameters ($A_\phi = 1$, $r = 0.04$), shown in Fig. 5.1. In Fig. 5.2 we show the mean and standard deviation over posterior samples drawn using the two methods, while in Fig. 5.3 we show the mean power spectrum of the samples. Additionally, we show the mean power spectrum of the reconstruction residual of samples ($\kappa_{\text{samples}} - \kappa_{\text{true}}$) in dashed lines, which quantifies the signal-to-noise ratio of the reconstruction on different scales. We see that for this configuration the reconstruction is signal-dominated up to $\ell \approx 800$. As a consequence, on small, noise-dominated scales the power spectrum of samples becomes increasingly prior driven, and the mean power spectrum of samples reduces to the theory spectrum, as observed in the bottom panel. We can further investigate the quality of our samples by comparing the correlation coefficient between samples and the true lensing convergence map as a function of scale:

$$\mathcal{C}_\ell = \frac{C_\ell^{\kappa_{\text{samples}}\kappa_{\text{true}}}}{\sqrt{C_\ell^{\kappa_{\text{true}}\kappa_{\text{true}}} C_\ell^{\kappa_{\text{samples}}\kappa_{\text{samples}}}}}, \tag{5.17}$$

shown in Figure 5.4. The various metrics demonstrate good agreement between the CMBLENSING.JL chain samples and SGM samples.

We conclude that our score-based generative model has learned an accurate ap-
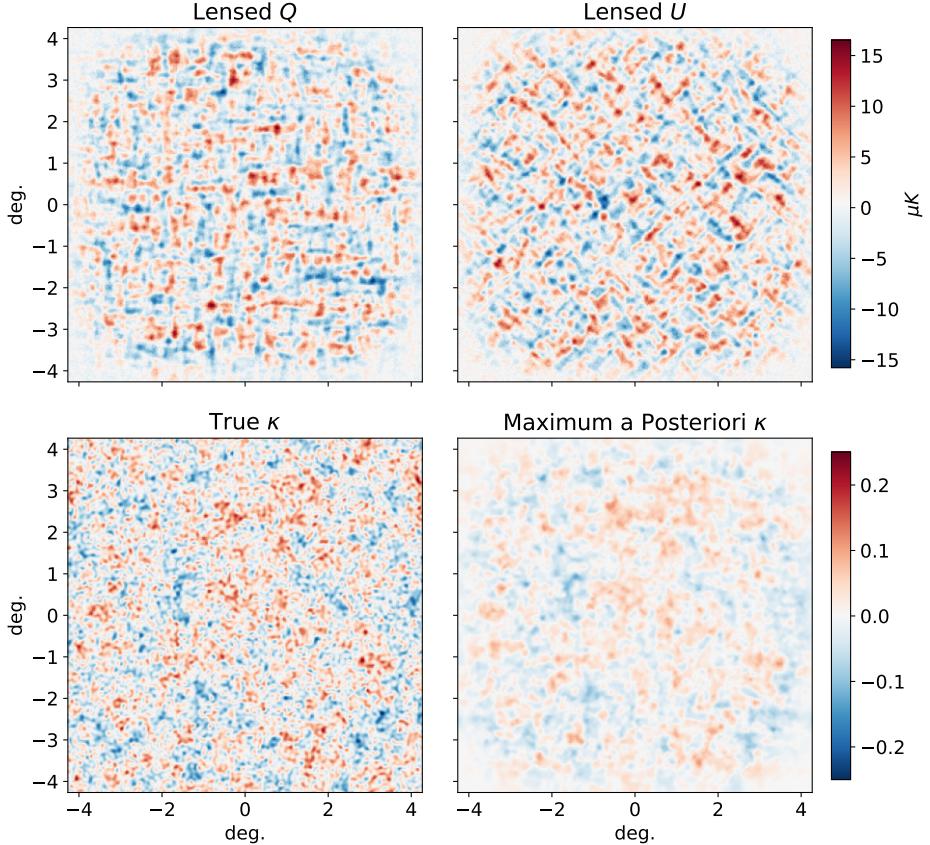
Figure 5.1: *Validation data used in the* CMBLENSING.JL *benchmark following the configuration presented in Table 5.1.* **Top***: the lensed* (Q, U) *observation.* **Bottom left***: the true lensing convergence map* κ *that lensed the* (Q, U) *observation.* **Bottom right***: the maximum a posteriori (MAP) estimate of the lensing convergence given the* (Q, U) *observation, using the optimization algorithm of* CMBLENSING.JL.

proximation of the lensing posterior, and can therefore be used to draw samples of the reconstructed lensing convergence. Our SGM model has the additional advantage that samples are uncorrelated, resulting in a more effective sampling of large angular scales compared to the CMBLENSING.JL chains, that struggle to converge here for masked data. In fact, for the CMBLENSING.JL chain presented here, the largest scale has not yet converged, explaining the large discrepancy of the largest scale bandpower with our SGM result, which more closely follows the truth.
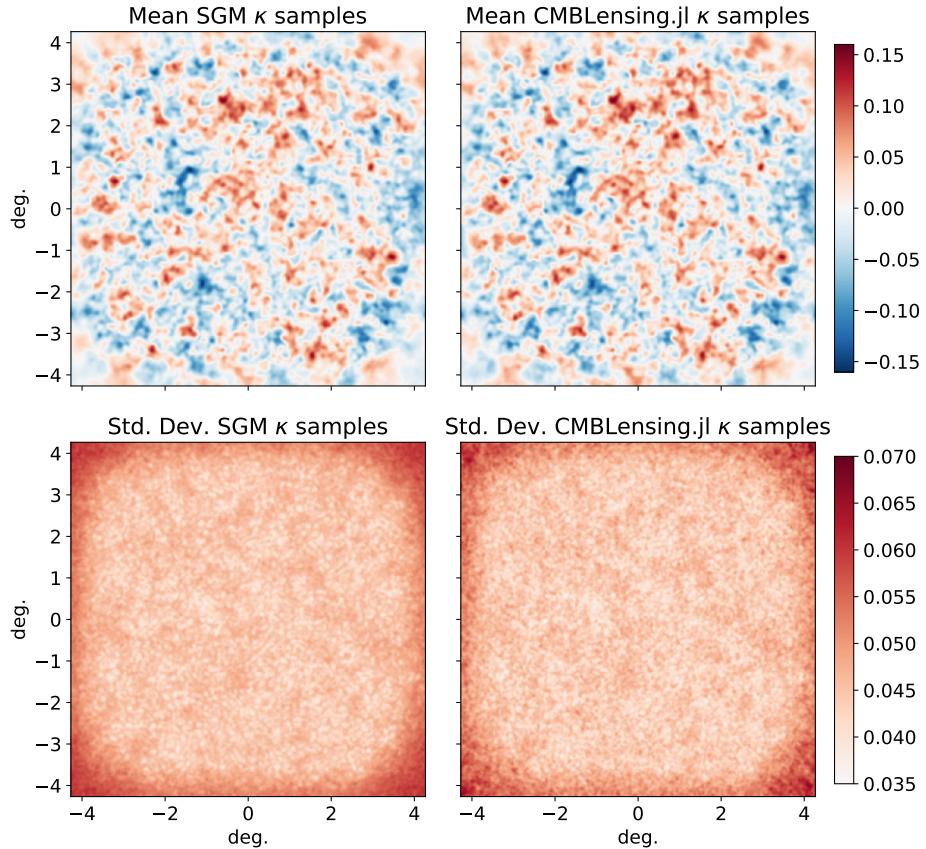
Figure 5.2: *Results for the validation sample of the* CMBLENSING.JL *benchmark.* **Top***: mean over 2048 posterior samples from the SGM (left) and the* CMBLENSING.JL *chain (right).* **Bottom***: standard deviation of posterior samples from 1024 diffusion samples (left) and the* CMBLENSING.JL *chain (right).*

5.6 Non-Gaussian Lensing reconstruction

Contrary to the MAW approach based on an explicitly specified posterior, our approach is based on a posterior that is implicitly specified through the training data. Our generative model therefore allows for a straightforward extension to non-Gaussian lensing potentials; we simply need to train our model on data consisting of non-Gaussian lensing maps and lensed CMB data. To demonstrate this, we train it using a set of simulated non-Gaussian lensing potentials provided by the authors of Ref. [61]. This dataset consists of 108 high-resolution full-sky lensing convergence
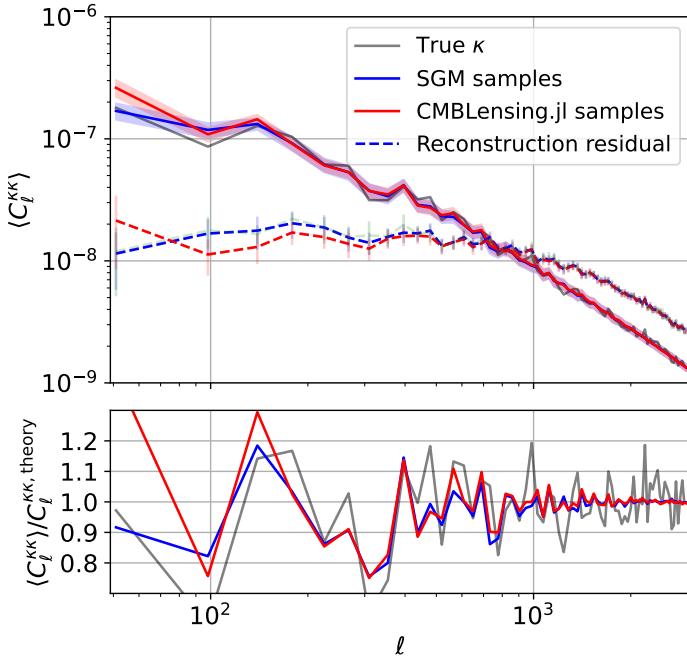
Figure 5.3: *Top: Mean power spectrum of SGM samples (blue),* CMBLENSING.JL *samples (red), and the true validation κ-map (gray). The reconstruction noise curves give the mean power spectrum of reconstruction residuals ($\kappa_{\mathrm{sample}} - \kappa_{\mathrm{true}}$). The shaded regions and error bars demark the 1σ spread of the power spectra. Bottom: mean power spectrum of samples divided by the theory power spectrum (at $A_\phi = 1$).*

maps generated by ray tracing dark matter N-body simulations.

We use 100 of their full-sky κ maps to lens 100 independent full-sky (Q, U)-maps using the lensing algorithm implemented in the `lenspyx` code [62]. We generate 256 flat-sky observations (and the corresponding patches of the lensing convergence) per full-sky map, with the same properties as before (i.e. Table 5.1). These 25600 flat-sky patches serve as training data for our model, while one of the remaining 8 full-sky maps is used to generate independent validation data. We keep cosmological parameters fixed ($A_\phi = 1$, $r = 0$) for all data for simplicity, but in principle, r can be varied in the training data as in the previous section.

Once trained, we draw 4096 samples from our model for a validation case, the results of which are shown in Figure 5.5. Both the samples and the mean over poste-
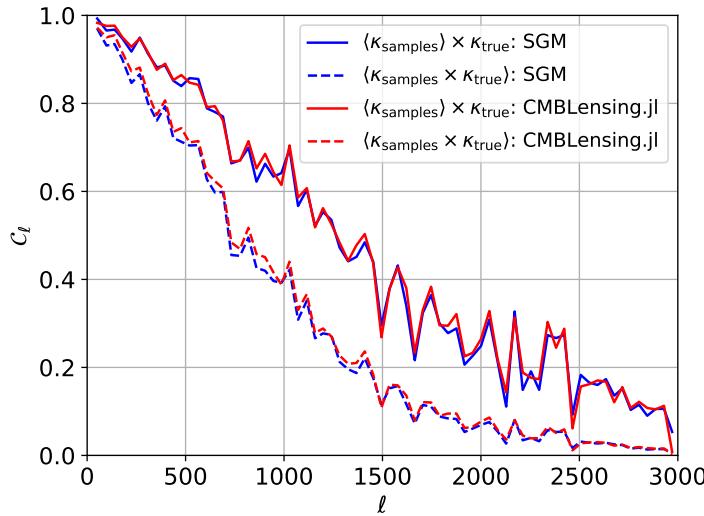
Figure 5.4: *Cross-correlations (Eq. (5.17)) of validation samples with the true lensing conver-gence map. Solid lines denote cross-correlation of the mean of posterior samples from the SGM (blue) and* CMBLENSING.JL *chain (red). Dashed lines denote the mean of individual cross-correlations of samples.*

rior samples show a clear correlation with the true non-Gaussian κ map. The power spectrum and cross-correlation coefficient for the non-Gaussian validation case are shown in Figure 5.6 and Figure 5.7.

Different from the Gaussian case of the previous section, the non-Gaussian lens-ing convergence maps contain information beyond the power spectrum. Such non-Gaussian statistics of the lensing field have been studied previously in [63], in partic-ular the one-point probability distribution function and peak counts. Additionally, we can study higher-order correlation functions, such as the bispectrum [40].

In Figure 5.8 we show the one-point PDF of validation κ samples, and that of both Gaussian and Non-Gaussian κ maps, clearly showing that our model produces one-point statistics that are consistent with the mean over many non-Gaussian maps, but slightly shifted towards the statistics of the true κ map.

Next, we investigate the bispectrum of the samples. We compute the bispectrum using 17 bins with a width of four times the largest mode, $\ell_{\min} \approx 43$, of the flat-sky patches, resulting in 597 triangle configurations up to $\ell_{\max} \approx 2911$. Figure 5.9a shows
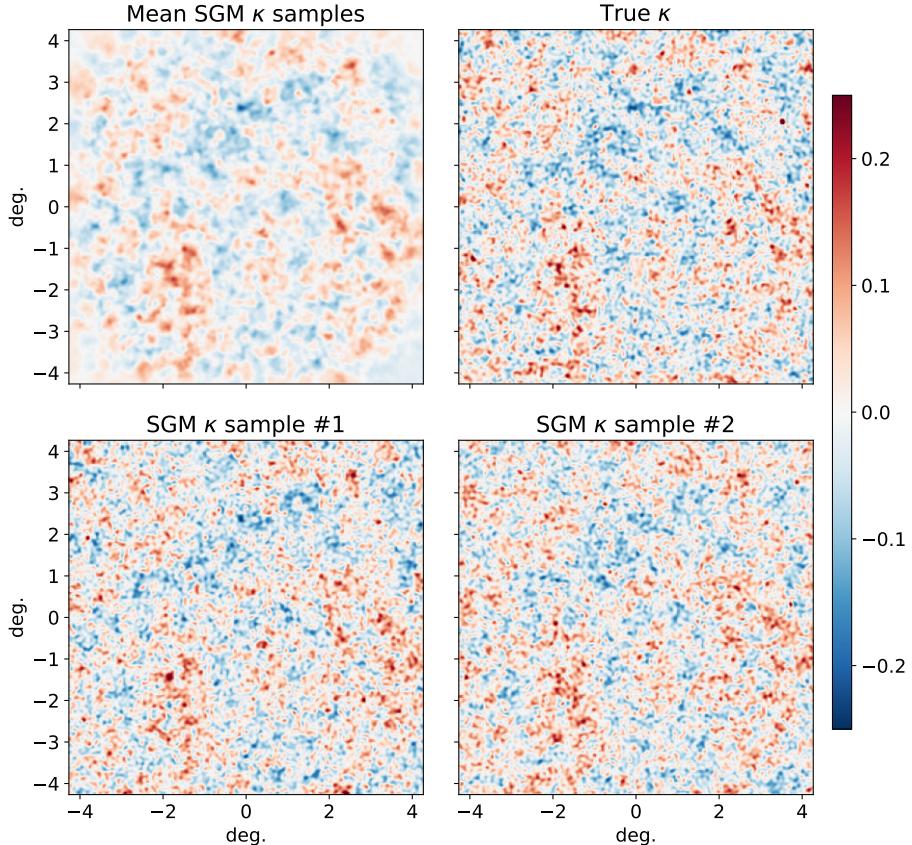
Figure 5.5: *Results for the validation case in the N-body non-Gaussian setup. **Top left:** mean over 4096 posterior samples. **Top right:** the true lensing convergence. **Bottom:** two independent samples of the convergence map.*

the mean bispectrum of samples and many N-body non-Gaussian maps. We plot the triangle configurations logarithmically to emphasize triangles on large scales, where the reconstruction is signal-dominated (as in Figure 5.6). On large scales (low triangle index), the mean bispectrum of samples (blue) follows the bispectrum of the true lensing convergence map (gray) more closely than it follows the theory bispectrum of non-Gaussian maps (green). On smaller scales, where the reconstruction is noise-dominated, the bispectrum of samples more closely follows the theory bispectrum (i.e. falls back to the learned non-Gaussian prior). These results show that our model has indeed learned a non-Gaussian posterior and prior and generates accurate samples of the non-Gaussian lensing convergence map.
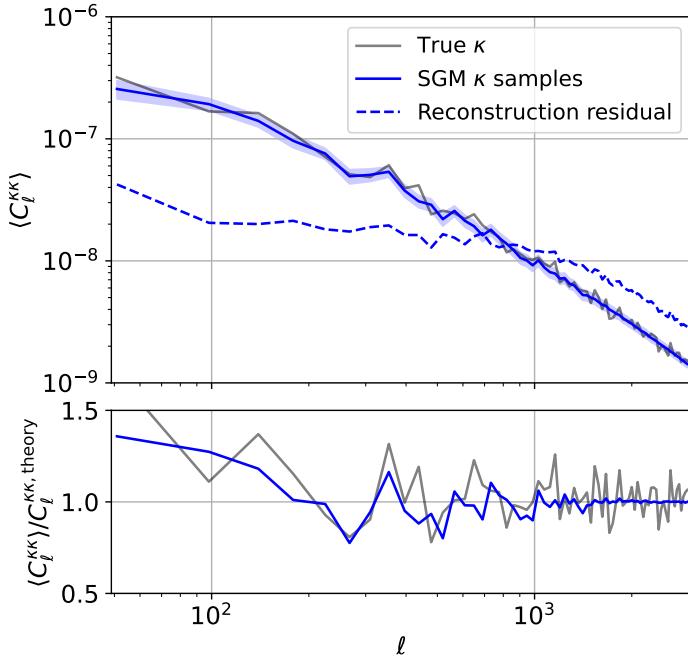
Figure 5.6: *Top:* *Mean power spectrum of SGM samples for the N-body validation case presented in Figure 5.5. The reconstruction noise curve gives the mean power spectrum of reconstruction residuals* $\kappa_{\mathrm{sample}} - \kappa_{\mathrm{true}}$. *Bottom:* *mean power spectrum of samples divided by the theory power spectrum (at $A_\phi = 1$).*

To demonstrate the error induced by drawing samples from the less informed posterior (with Gaussian lensing prior) we show the bispectrum for samples of the validation case drawn from a model that was trained using only Gaussian lensing maps (with the same power spectrum as the non-Gaussian N-body lensing maps). In Figure 5.9b we show the mean bispectrum of 4096 validation samples from this Gaussian model. We expect such a model to be able to capture non-Gaussian information only on highly signal-dominated regions where the posterior is not dominated by the Gaussian prior (that is implicitly learned from the Gaussian training data). Indeed the top panel shows that on the largest, signal-dominated scales, the model produces a bispectrum that follows the true validation map and is similar to that of the SGM trained on non-Gaussian maps. However, on smaller, noise-dominated scales the bispectrum vanishes, as the model falls back on the internally learned Gaussian prior, unlike our non-Gaussian model. Furthermore, in the bot-
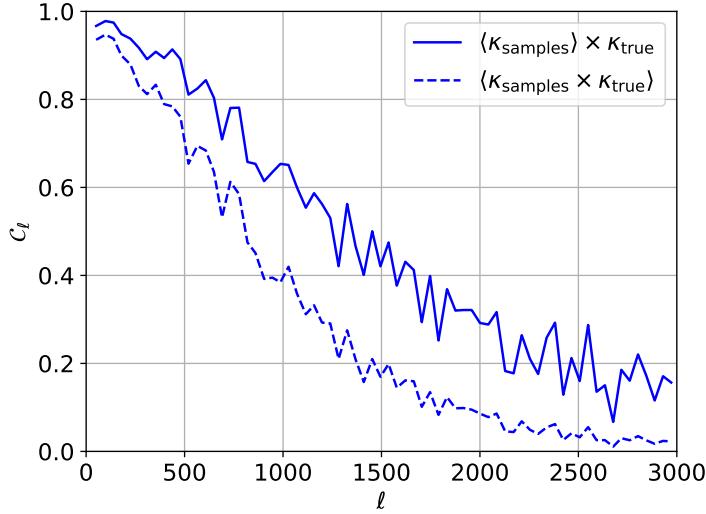
Figure 5.7: *Cross-correlations of validation samples with the true lensing convergence map for the non-Gaussian case. Solid lines denote the cross-correlation of the mean of posterior samples from the SGM. Dashed lines denote the mean of individual cross-correlations of samples.*

tom panel, we see that the mean bispectrum over 4096 samples of different lensing realizations, does not closely follow the theory bispectrum.

5.7   Conclusion & Outlook

In this work, we demonstrated the use of score-based generative models in CMB lensing reconstruction tasks. We have shown that such models can effectively learn the conditional posterior distribution of the lensing convergence map given observed CMB data, by training on corresponding pairs of these. The trained model can be used to rapidly draw lensing posterior samples. For the case of Gaussian lensing convergence maps, we have validated our methodology by comparing the results of our model against the established Hamiltonian Monte Carlo methods of Millea, Anderes and Wandelt [45].

Since our model learns the posterior distribution entirely from simulated data, and the prior and likelihood are modeled through the processes encoded in the simulations, it can be used in situations where the analytic form of the posterior is not known, such as in the case of non-Gaussian lensing convergence maps. Using re-
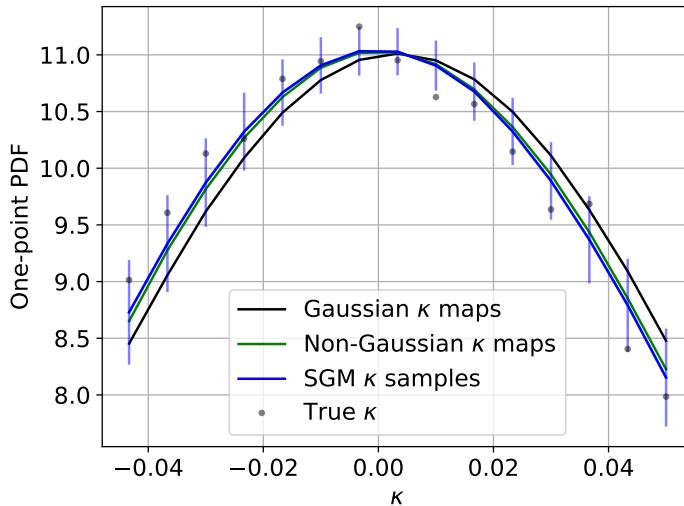
Figure 5.8: *One-point probability distribution function of κ, using 15 linear bins between $\kappa \in (-0.05, 0.05)$. In green, we show the PDF of a set of the N-body non-Gaussian κ maps. In black, we show the PDF of a set of Gaussian κ maps generated with the same power spectrum as the non-Gaussian maps. In blue, we show the PDF of a set of samples using our generative model, and the error bars denote its standard deviation per bin. Gray dots denote the one-point PDF of the true validation κ map.*

alistic non-Gaussian lensing maps obtained by ray-tracing N-body simulations, we demonstrate that our model can indeed be trained to yield samples with the correct non-Gaussian statistics (i.e. one-point probability distribution and bispectrum). As we have shown, the resulting samples of our model can be used to estimate the CMB lensing bispectrum, which is a difficult task using traditional estimators such as the quadratic estimator, because of large noise biases [41].

Another important advantage of our model as compared to the HMC method is speed and efficiency. Our model, with neural network configuration as described in Appendix 5.A and using 1000 denoising steps, generates a single sample in 12 seconds, which can be vectorized up to 128 simultaneous samples in about 225 seconds (~ 1.75 seconds per sample) on an A100 GPU, and can be easily parallelized over multiple GPUs. Significant speed improvements can be expected from optimizing neural network complexity and noise schedule (i.e. number of denoising steps), which is beyond the scope of this chapter. Notably, the samples of our model are completely uncorrelated. Meanwhile, drawing a single sample with CMBLENS-ING.JL (in the configuration used in this chapter, varying only r) takes ~ 1.6 seconds
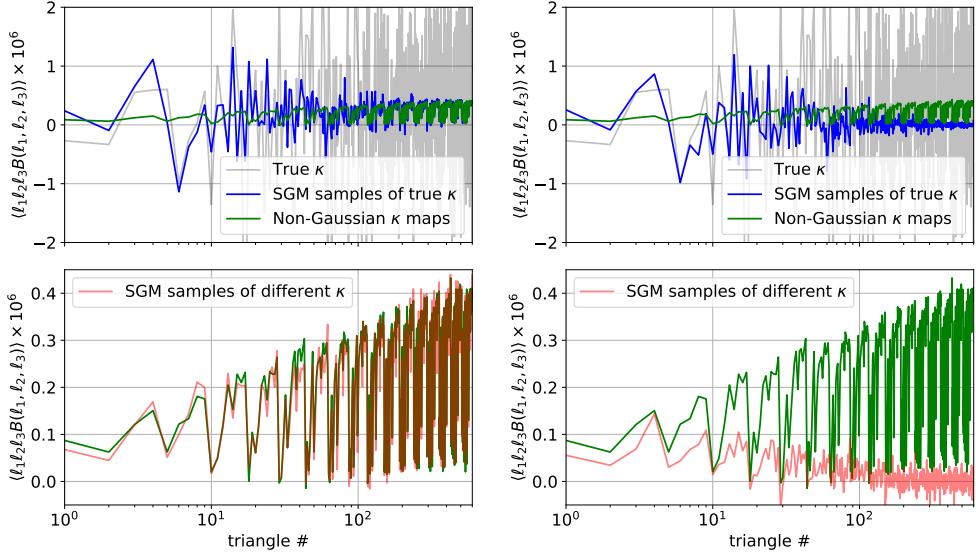
Figure 5.9: *Bispectrum results for an SGM trained on non-Gaussian (left) or Gaussian (right) lensing data.* **Top:** *bispectrum of the true validation lensing convergence κ map (gray), and mean bispectrum of many different non-Gaussian κ maps (green), and SGM samples of the true κ map (blue).* **Bottom:** *mean bispectrum of SGM samples of many different κ maps (red). The green line is an approximation to the theory bispectrum.*

on an A100 GPU. However, HMC samples are correlated. Taking the autocorrelation length of the bandpower on the least efficiently sampled scale of the chain (which is found to be about 290), shows that the effective independent sample size of the chain is more than two orders of magnitude smaller. It should be noted that improving the HMC analysis with suitable approximations used in real lensing analyses, is likely to again yield an order of magnitude improvement in speed for the configuration considered in this work [46, 47]. Finally, for larger maps, the HMC analysis requires more samples to converge, thus scaling inefficiently with problem size. For our model, besides the obvious increase in training and sampling time due to the larger data size, there is no such chain convergence issue, thus promising more efficient scaling to larger problems. We conclude that generative models allow for a more efficient approach to Bayesian CMB lensing reconstruction, which translates to faster convergence of analysis pipelines.

The methods described in this paper, as well as those in [34, 43, 45], are so far limited to only using a small and flat patch of the sky at a time. Although adaptation of our methods to curved-sky geometry using spherical convolutional neural networks

(e.g. [64, 65]) is in principle straightforward, such an analysis is likely to be computationally prohibitive at present. Nevertheless, we can expect ongoing advancements in machine learning and computational technology to make such analyses more feasible. Additionally, it is worth noting that validation of conventional methods, such as the quadratic or iterative estimator, also requires many evaluations on simulated data to determine the uncertainty. In our methods, as well as the HMC method, this is inherently included in the analysis.

Looking ahead, ideally, the model would include the amplitude of the CMB lensing power spectrum as well. As mentioned earlier, we have been unsuccessful at this so far. Nevertheless, we are optimistic that with an appropriate modification of the model or algorithm, this can be overcome. We share some of our thoughts on this issue in Appendix 5.B. Alternatively, sampling the cosmological parameters can also be left as a seperate Gibbs-like steps using another simulation-based inference pipeline conditioned on the data and our model's lensing samples. We leave an investigation into this for future work.

Though in this work we have focused on the non-Gaussianity of the lensing potential, our model's flexibility straightforwardly allows for other extensions that would similarly complicate the posterior evaluation, such as the inclusion of inhomogeneous and anisotropic noise in the data [66]. Once again, this would require only a modification of the simulated data that the model is trained on. Along these lines, obvious directions for future work are to assess biases due to foregrounds [22, 67, 68] and correlated point-source masks [69] that complicate the reconstruction of the lensing potential, and reconstruction of lensing curl modes [70–74].

The results of this work pave the way for more efficient, accurate, and realistic lensing reconstruction. The resulting samples of our model can be used in subsequent analyses of the CMB, or as an unbiased tracer of the total matter density distribution while being able to appropriately propagate errors due to uncertainty in the lensing reconstruction.

Our model and the code used to generate the results of this chapter is publicly available at `https://github.com/tsfloss/CMBLensingDiffusion`

5.A   Neural Network Specifics

The U-Net used in our score-based generative model employs double convolution blocks, each of which consists of a 3×3 convolution with 64 channels, a sinu-

soidal time embedding, group normalization, and SiLU activation, followed (preceded) by a Fourier pooling (upsampling) layer in the encoding (decoding) process. The Fourier pooling layer performs the downsampling of the image in the Fourier domain, conserving the data's frequency content more accurately than typical pooling or stride convolutions. Similarly, the Fourier upsampling (or interpolation) layer upsamples the data in the Fourier domain, by filling up the modes of a higher-resolution image with the modes of the lower-resolution image, again exactly preserving the frequency content. We furthermore employ residual connections (rescaled with $\sqrt{2}$) and skip-connections. At the network bottleneck, the data has a size of 8^2 pixels. The network input is a channel-wise concatenation of the noised image $\mathbf{x}_i$ and conditioning data $\mathbf{y}$ [60]. The model is trained using the AdamW optimizer with gradient norms clipped to 1, and a batch size of 128. We first train with a learning rate of 0.0002, followed by a learning rate of 10^{-5} to ensure convergence. Additionally, we use exponential moving average (EMA) weights, with a decay rate of 0.9999, to further stabilize the model during inference. Each model is trained for a total of around 200000 steps, taking around 16 hours on two A100 GPUs. Our model is implemented using JAX [75].

5.B Varying the lensing amplitude

In Section 5.5, we only varied the tensor-to-scalar ratio r. Ideally, we would like to be able to marginalize over the amplitude A_ϕ of the lensing power spectrum too. However, we found that the diffusion algorithm as is, is unable to yield samples with the correct amplitude on small scales in this case, even though the phases of the samples are accurate (as quantified by the cross-correlation coefficient in equation (5.17), that does not contain the amplitude). We find that distinct models converge to different results. An example of this is visualized in Figure 5.10, where we show results of the mean power spectrum of samples for three different realizations and parameters, from two models (solid and dotted) trained on data that simultaneously vary A_ϕ and r, compared to those of CMBLENSING.JL chains. These two models have also been trained for a significantly longer time than the models presented in the main text, to ensure convergence of their EMA weights.

Although we are uncertain as to what exactly causes this issue, we think that it could be related to the nature of the diffusion sampling algorithm. The algorithm assumes that the signal-to-noise ratio of the noised image at different timesteps follows the noise schedule $\sigma(t)$ in equation (5.14). However, when varying the amplitude of the underlying data, at a fixed time this ratio varies significantly among the data, especially at lower noise levels. Meanwhile, during the sampling process, at every

denoising step, the network makes a guess for the amplitude that will not be perfect. Thus, at the next timestep, there is a mismatch between the actual signal-to-noise ratio and that expected by the model. The accumulation of many of these mismatches could the wrong amplitude, especially on noise-dominated scales (where the value of the amplitude is most relevant), even though the phase structure is accurately resolved. From this perspective, a blind denoising approach, such as that presented in [76], might prove useful.
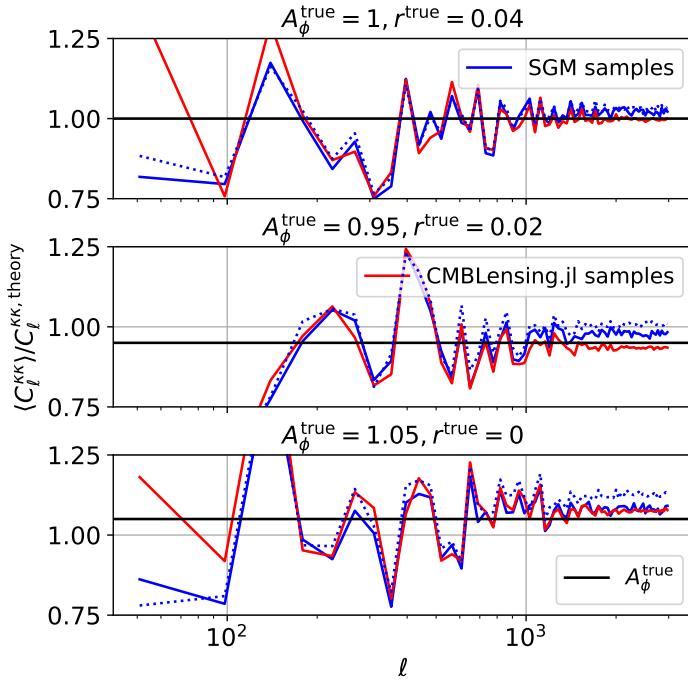

Figure 5.10: *Mean power spectrum, normalized by the theory power spectrum at $A_\phi = 1$, of samples from different models (solid and dotted blue) and* CMBLENSING.JL *chains (red). for three different realizations and parameters.*

Bibliography

[1] S. Cole and G. Efstathiou, "Gravitational lensing of fluctuations in the microwave background radiation," *Monthly Notices of the Royal Astronomical Society* **239**, pp. 195–200, July 1989.

[2] E. V. Linder, "Analysis of gravitationally lensed microwave background anisotropies," *Monthly Notices of the Royal Astronomical Society* **243**, pp. 353–361, Apr. 1990.

[3] U. Seljak, "Gravitational lensing effect on cosmic microwave background anisotropies: A Power spectrum approach," *Astrophys. J.* **463**, p. 1, 1996.

[4] R. B. Metcalf and J. Silk, "Gravitational magnification of the cosmic microwave background," *Astrophys. J.* **489**, pp. 1–6, 1997.

[5] A. Lewis and A. Challinor, "Weak gravitational lensing of the CMB," *Phys. Rept.* **429**, pp. 1–65, 2006.

[6] U. Seljak and C. M. Hirata, "Gravitational lensing as a contaminant of the gravity wave signal in CMB," *Phys. Rev. D* **69**, p. 043005, 2004.

[7] D. Babich and M. Zaldarriaga, "Primordial bispectrum information from CMB polarization," *Phys. Rev. D* **70**, p. 083005, 2004.

[8] W. R. Coulton, P. D. Meerburg, D. G. Baker, S. Hotinli, A. J. Duivenvoorden, and A. van Engelen, "Minimizing gravitational lensing contributions to the primordial bispectrum covariance," *Phys. Rev. D* **101**(12), p. 123504, 2020.

[9] W. Coulton, A. Miranthis, and A. Challinor, "Biases to primordial non-Gaussianity measurements from CMB secondary anisotropies," *Mon. Not. Roy. Astron. Soc.* **523**, pp. 825–848, 2023.

[10] A. Achúcarro *et al.*, "Inflation: Theory and Observations," 3 2022.

[11] D. Green, J. Meyers, and A. van Engelen, "CMB Delensing Beyond the B Modes," *JCAP* **12**, p. 005, 2017.

[12] S. C. Hotinli, J. Meyers, C. Trendafilova, D. Green, and A. van Engelen, "The benefits of CMB delensing," *JCAP* **04**(04), p. 020, 2022.

[13] C. Trendafilova, S. C. Hotinli, and J. Meyers, "Improving Constraints on Inflation with CMB Delensing," 12 2023.

[14] J. Ange and J. Meyers, "Improving constraints on models addressing the Hubble tension with CMB delensing," *JCAP* **10**, p. 045, 2023.

[15] K. N. Abazajian *et al.*, "Neutrino Physics from the Cosmic Microwave Background and Large Scale Structure," *Astropart. Phys.* **63**, pp. 66–80, 2015.

[16] U. Seljak, "Extracting primordial non-gaussianity without cosmic variance," *Phys. Rev. Lett.* **102**, p. 021302, 2009.

[17] M. Schmittfull and U. Seljak, "Parameter constraints from cross-correlation of CMB lensing with galaxy clustering," *Phys. Rev. D* **97**(12), p. 123540, 2018.

[18] M. Zaldarriaga and U. Seljak, "Reconstructing projected matter density from cosmic microwave background," *Phys. Rev. D* **59**, p. 123507, 1999.

[19] W. Hu and T. Okamoto, "Mass reconstruction with cmb polarization," *Astrophys. J.* **574**, pp. 566–574, 2002.

[20] T. Okamoto and W. Hu, "CMB lensing reconstruction on the full sky," *Phys. Rev. D* **67**, p. 083002, 2003.

[21] S. Das *et al.*, "Detection of the Power Spectrum of Cosmic Microwave Background Lensing by the Atacama Cosmology Telescope," *Phys. Rev. Lett.* **107**, p. 021301, 2011.

[22] A. van Engelen *et al.*, "A measurement of gravitational lensing of the microwave background using South Pole Telescope data," *Astrophys. J.* **756**, p. 142, 2012.

[23] P. A. R. Ade *et al.*, "Measurement of the Cosmic Microwave Background Polarization Lensing Power Spectrum with the POLARBEAR experiment," *Phys. Rev. Lett.* **113**, p. 021301, 2014.

[24] K. T. Story *et al.*, "A Measurement of the Cosmic Microwave Background Gravitational Lensing Potential from 100 Square Degrees of SPTpol Data," *Astrophys. J.* **810**(1), p. 50, 2015.

[25] P. A. R. Ade *et al.*, "Planck 2013 results. XVII. Gravitational lensing by large-scale structure," *Astron. Astrophys.* **571**, p. A17, 2014.

[26] P. A. R. Ade *et al.*, "Planck 2015 results. XV. Gravitational lensing," *Astron. Astrophys.* **594**, p. A15, 2016.

[27] B. A. Benson *et al.*, "SPT-3G: A Next-Generation Cosmic Microwave Background Polarization Experiment on the South Pole Telescope," *Proc. SPIE Int. Soc. Opt. Eng.* **9153**, p. 91531P, 2014.

[28] K. Prabhu *et al.*, "Testing the ΛCDM Cosmological Model with Forthcoming Measurements of the Cosmic Microwave Background with SPT-3G," 3 2024.

[29] P. Ade *et al.*, "The Simons Observatory: Science goals and forecasts," *JCAP* **02**, p. 056, 2019.

[30] K. N. Abazajian *et al.*, "CMB-S4 Science Book, First Edition," 10 2016.

[31] C. M. Hirata and U. Seljak, "Analyzing weak lensing of the cosmic microwave background using the likelihood function," *Phys. Rev. D* **67**, p. 043001, 2003.

5

[32] C. M. Hirata and U. Seljak, "Reconstruction of lensing from the cosmic microwave background polarization," *Phys. Rev. D* **68**, p. 083002, 2003.

[33] J. Carron and A. Lewis, "Maximum a posteriori CMB lensing reconstruction," *Phys. Rev. D* **96**(6), p. 063510, 2017.

[34] M. Millea, E. Anderes, and B. D. Wandelt, "Bayesian delensing of CMB temperature and polarization," *Phys. Rev. D* **100**(2), p. 023509, 2019.

[35] B. Hadzhiyska, B. D. Sherwin, M. Madhavacheril, and S. Ferraro, "Improving Small-Scale CMB Lensing Reconstruction," *Phys. Rev. D* **100**(2), p. 023547, 2019.

[36] J. a. Caldeira, W. L. K. Wu, B. Nord, C. Avestruz, S. Trivedi, and K. T. Story, "DeepCMB: Lensing Reconstruction of the Cosmic Microwave Background with Deep Neural Networks," *Astron. Comput.* **28**, p. 100307, 2019.

[37] E. Guzman and J. Meyers, "Reconstructing patchy reionization with deep learning," *Phys. Rev. D* **104**(4), p. 043529, 2021.

[38] P. Li, I. I. Onur, S. Dodelson, and S. Chaudhari, "High-Resolution CMB Lensing Reconstruction with Deep Learning," 5 2022.

[39] Y.-P. Yan, G.-J. Wang, S.-Y. Li, Y.-J. Yan, and J.-Q. Xia, "Lensing Reconstruction from the Cosmic Microwave Background Polarization with Machine Learning," *Astrophys. J.* **952**(1), p. 15, 2023.

[40] T. Namikawa, "CMB Lensing Bispectrum from Nonlinear Growth of the Large Scale Structure," *Phys. Rev. D* **93**(12), p. 121301, 2016.

[41] A. Kalaja, G. Orlando, A. Bowkis, A. Challinor, P. D. Meerburg, and T. Namikawa, "The reconstructed CMB lensing bispectrum," *JCAP* **04**, p. 041, 2023.

[42] T. Namikawa, D. Hanson, and R. Takahashi, "Bias-hardened cmb lensing," *Monthly Notices of the Royal Astronomical Society* **431**, p. 609–620, Feb. 2013.

[43] M. Millea, E. Anderes, and B. D. Wandelt, "Sampling-based inference of the primordial CMB and gravitational lensing," *Phys. Rev. D* **102**(12), p. 123542, 2020.

[44] E. Anderes, B. Wandelt, and G. Lavaux, "Bayesian inference of CMB gravitational lensing," *Astrophys. J.* **808**(2), p. 152, 2015.

[45] M. Millea *et al.*, "Optimal Cosmic Microwave Background Lensing Reconstruction and Parameter Estimation with SPTpol Data," *Astrophys. J.* **922**(2), p. 259, 2021.

[46] M. Millea, "Improved marginal unbiased score expansion (muse) via implicit differentiation," 2022.

[47] M. Millea and U. Seljak, "Marginal unbiased score expansion and application to CMB lensing," *Phys. Rev. D* **105**(10), p. 103531, 2022.

[48] K. Cranmer, J. Brehmer, and G. Louppe, "The frontier of simulation-based inference," *Proc. Nat. Acad. Sci.* **117**(48), pp. 30055–30062, 2020.

[49] J. Alsing, T. Charnock, S. Feeney, and B. Wandelt, "Fast likelihood-free cosmology with neural density estimators and active learning," *Monthly Notices of the Royal Astronomical Society* **488**, pp. 4440–4458, Sept. 2019.

[50] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-Based Generative Modeling through Stochastic Differential Equations," *arXiv e-prints* , p. arXiv:2011.13456, Nov. 2020.

[51] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," 2020.

[52] R. Legin, M. Ho, P. Lemos, L. Perreault-Levasseur, S. Ho, Y. Hezaveh, and B. Wandelt, "Posterior sampling of the initial conditions of the universe from non-linear large scale structures using score-based generative models," *Mon. Not. Roy. Astron. Soc.* **527**(1), pp. L173–L178, 2023.

[53] K. Karchev, N. Anau Montel, A. Coogan, and C. Weniger, "Strong-Lensing Source Reconstruction with Denoising Diffusion Restoration Models," in *36th Conference on Neural Information Processing Systems: Workshop on Machine Learning and the Physical Sciences*, 11 2022.

[54] V. Ono, C. F. Park, N. Mudur, Y. Ni, C. Cuesta-Lazaro, and F. Villaescusa-Navarro, "Debiasing with Diffusion: Probabilistic reconstruction of Dark Matter fields from galaxies with CAMELS," 3 2024.

[55] A. Schanz, F. List, and O. Hahn, "Stochastic Super-resolution of Cosmological Simulations with Denoising Diffusion Models," 10 2023.

[56] A. Rouhiainen, M. Gira, M. Münchmeyer, K. Lee, and G. Shiu, "Super-Resolution Emulation of Large Cosmological Fields with a 3D Conditional Diffusion Model," 11 2023.

[57] D. Heurtel-Depeiges, B. Burkhart, R. Ohana, and B. R.-S. Blancard, "Removing Dust from CMB Observations with Diffusion Models," 10 2023.

[58] A. Challinor and G. Chon, "Geometry of weak lensing of CMB polarization," *Phys. Rev. D* **66**, p. 127301, 2002.

[59] B. D. Anderson, "Reverse-time diffusion equation models," *Stochastic Processes and their Applications* **12**(3), pp. 313–326, 1982.

[60] G. Batzolis, J. Stanczuk, C.-B. Schönlieb, and C. Etmann, "Conditional image generation with score-based diffusion models," 2021.

[61] R. Takahashi, T. Hamana, M. Shirasaki, T. Namikawa, T. Nishimichi, K. Osato, and K. Shiroyama, "Full-sky Gravitational Lensing Simulation for Large-area Galaxy Surveys and Cosmic Microwave Background Experiments," *Astrophys. J.* **850**(1), p. 24, 2017.

5

[62] M. Reinecke, S. Belkner, and J. Carron, "Improved cosmic microwave background (de-)lensing using general spherical harmonic transforms," *Astron. Astrophys.* **678**, p. A165, 2023.

[63] J. Liu, J. C. Hill, B. D. Sherwin, A. Petri, V. Böhm, and Z. Haiman, "CMB lensing beyond the power spectrum: Cosmological constraints from the one-point probability distribution function and peak counts," *Phys. Rev. D* **94**(10), p. 103501, 2016.

[64] M. Defferrard, M. Milani, F. Gusset, and N. Perraudin, "DeepSphere: a graph-based spherical CNN," *arXiv e-prints* , p. arXiv:2012.15000, Dec. 2020.

[65] C. Esteves, J.-J. Slotine, and A. Makadia, "Scaling Spherical CNNs," *arXiv e-prints* , p. arXiv:2306.05420, June 2023.

[66] D. Hanson, G. Rocha, and K. GÃ³rski, "Lensing reconstruction from planck sky maps: inhomogeneous noise: Lensing reconstruction from planck sky maps," *Monthly Notices of the Royal Astronomical Society* **400**, p. 2169–2173, Dec. 2009.

[67] S. J. Osborne, D. Hanson, and O. Doré, "Extragalactic Foreground Contamination in Temperature-based CMB Lens Reconstruction," *JCAP* **03**, p. 024, 2014.

[68] N. MacCrann *et al.*, "The Atacama Cosmology Telescope: Mitigating the impact of extragalactic foregrounds for the DR6 CMB lensing analysis," 4 2023.

[69] M. Lembo, G. Fabbian, J. Carron, and A. Lewis, "CMB lensing reconstruction biases from masking extragalactic sources," *Phys. Rev. D* **106**(2), p. 023525, 2022.

[70] A. Cooray, M. Kamionkowski, and R. R. Caldwell, "Cosmic shear of the microwave background: The Curl diagnostic," *Phys. Rev. D* **71**, p. 123527, 2005.

[71] T. Namikawa, D. Yamauchi, and A. Taruya, "Full-sky lensing reconstruction of gradient and curl modes from CMB maps," *JCAP* **01**, p. 007, 2012.

[72] T. Namikawa, D. Yamauchi, and A. Taruya, "Constraining cosmic string parameters with curl mode of CMB lensing," *Phys. Rev. D* **88**(8), p. 083525, 2013.

[73] G. Pratten and A. Lewis, "Impact of post-Born lensing on the CMB," *JCAP* **08**, p. 047, 2016.

[74] M. Robertson and A. Lewis, "How to detect lensing rotation," *JCAP* **08**, p. 048, 2023.

[75] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, "JAX: composable transformations of Python+NumPy programs," 2018.

[76] D. Heurtel-Depeiges, C. C. Margossian, R. Ohana, and B. R.-S. Blancard, "Listening to the Noise: Blind Denoising with Gibbs Diffusion," 2 2024.

5

Chapter 6

Non-linear Electrodynamics from Massive Gravity

Abstract

As a counterpart to the four-fermion interaction, which describes massive vector exchange at low energies, we investigate the low-energy effective action of photons under exchange of a massive graviton. We show how integrating out a massive graviton leads to the most general duality-invariant vector interactions in 4D or, vice versa, how any such interactions have a natural interpretation within massive gravity. Moreover, we demonstrate how the special case of Born-Infeld theory arises from arguably the simplest graviton potential within ghost-free dRGT massive gravity.

6.1   Introduction

Effective field theories are ubiquitous throughout physics. Perhaps the most famous example is the low-energy description of β-decay in terms of four-fermion interactions [1]. With the development of the electroweak theory and the Higgs mechanism [2, 3], it became clear that at a more fundamental level, this physical process is mediated by the $W_\pm$ and Z_0 vector bosons. Fermi's theory then arises as a low-energy description of this process, valid below the electroweak scale of ~ 100 GeV, in terms of effective contact interactions for the fermions.

This top-down approach to effective field theories can be more broadly applied to any theory with a separation of scales between lighter and heavier degrees of freedom: at energies well below the heavy mass, there is insufficient energy available to excite the heavier modes. These can therefore be integrated out, resulting in effective interactions between the lighter degrees of freedom. This philosophy is independent of the nature of the heavier degrees of freedom and can also be applied to massive gravitons.

Massive gravity (MG), originally formulated with an eye towards the cosmological constant problem, is a highly non-trivial and interesting construction from a purely theoretical perspective. Remarkably, ensuring the absence of ghosts allows for two free parameters (besides the graviton mass) in 4D [4, 5]; in contrast to the electroweak theory, no Higgs-like mechanism is known to connect the massive and massless phases. Moreover, the parameter space of MG is subject to positivity bounds [6], with a very recent claim that these do not allow for a parametric separation of graviton mass well below the cut-off scale [7].

In this chapter, we will consider effective theories emerging from integrating out massive gravitons (and thus the opposite limit of [7]). For simplicity we will take the lighter degrees of freedom to be massless, and focus on the case of a Maxwell vector that only interacts gravitationally. The resulting low-energy description, with effective self-interactions of the Maxwell vector, spans an interesting set of theories: it consists of all non-linear completions that preserve the electromagnetic duality invariance of the original Maxwell theory[1] (in contrast to e.g. the Euler-Heisenberg theory [10] that follows from integrating out massive electrons at one-loop). All of these effective theories are expected to be fully compatible with positivity bounds; see Ref. [11, 12] for such bounds for $2 \to 2$ scattering.

---

[1]Duality invariance also plays an important role in supersymmetric theories, where e.g. the $E_{7(7)}$ duality invariance features prominently in the discussion of possible finiteness of $\mathcal{N} = 8$ supergravity [8, 9].

In the second half of this chapter, we will restrict ourselves to the ghost-free dRGT theory of massive gravity [4, 5]. We demonstrate how, starting from a particularly simple element of dRGT, one can generate the most important example of non-linear electrodynamics: Born-Infeld theory. This theory describes the world-volume of D-branes in string theory [13] and has numerous special amplitude properties, see e.g. [14, 15]. It is therefore interesting that BI arises as an effective description of graviton exchange. Moreover, Born-Infeld and gravity-coupled Maxwell theories both appear in the double copy framework of [16], as illustrated in the tetrahedron of [17]; it thus appears that moving to massive exchange particles corresponds to a particular operation in the double copy web of relations, similar to [14, 18].

We comment on further links and implications and future extensions in the concluding section. Amongst these is the analogon of our discussion in 2D space-time, where the idea of integrating out massive gravity was worked out in 2D in Ref. [19]. It was shown that coupling dRGT gravity to a generic field theory is equivalent to performing a $T\bar{T}$-deformation of said theory. Our work can thus be seen as the extension of this procedure to the case of conformally-coupled Maxwell theory in 4D, suggesting a close connection to $T\bar{T}$-deformations of this theory, as studied in Refs. [20–24].

Notation. We work in 4D spacetime throughout and employ matrix notation, with e.g. $F \equiv F_{\mu\nu}$ and $F^2 \equiv F_{\alpha\gamma}\eta^{\gamma\delta}F_{\delta\beta}$ (note the ordering of indices). Traces taken with the flat metric η are written using square brackets, e.g. $[F^2] = F_{\alpha\beta}\eta^{\beta\gamma}F_{\gamma\delta}\eta^{\delta\alpha}$. Angled brackets instead denote traces where all tensors are contracted with the full metric g, e.g. $\langle F^2 \rangle = F_{\alpha\beta}g^{\beta\gamma}F_{\gamma\delta}g^{\delta\alpha}$ (following the conventions of [25]).

6

6.2 Integrating out gravity

We start by integrating out the massive graviton perturbatively, obtaining the effective vector contact interactions order by order. Our set-up consists of a Maxwell field minimally coupled to massive gravity:

$$\mathcal{L}_{gF^2} + \mathcal{L}_g = \sqrt{-g}\left[\frac{1}{4}\langle F^2 \rangle + \frac{1}{2}R - \frac{1}{8}m^2 \sum_i V^{(i)}(h)\right]. \tag{6.1}$$

where $g = \eta + h$. The most general graviton potential terms are given at lowest order by [25]

$$\begin{aligned}
V^{(2)} &= \langle h^2 \rangle + b_2 \langle h \rangle^2, \\
V^{(3)} &= c_1 \langle h^3 \rangle + c_2 \langle h \rangle \langle h^2 \rangle + c_3 \langle h \rangle^3, \\
V^{(4)} &= d_1 \langle h^4 \rangle + d_2 \langle h^3 \rangle \langle h \rangle + d_3 \langle h^2 \rangle^2 + d_4 \langle h^2 \rangle \langle h \rangle^2 + d_5 \langle h \rangle^4.
\end{aligned} \tag{6.2}$$

At energies well below the mass scale of the graviton, we can neglect the Einstein-Hilbert term so the massive graviton becomes an auxiliary field and can be integrated out through its algebraic equation of motion[2],

$$\frac{1}{\sqrt{-g}}\frac{\delta\sqrt{-g}V(g)}{\delta g^{-1}} = T^{(\text{Maxwell})}, \tag{6.3}$$

in matrix notation.

Proceeding perturbatively, we expand the metric as $h = h^{(1)} + h^{(2)} + \ldots$ and solve the above equation of motion order by order, resulting in effective contact interactions for the vector field. At lowest order, the graviton field equation is solved by

$$h^{(1)} = -\frac{2}{m^2}F^2 + \frac{1}{2m^2}[F^2]\eta. \tag{6.4}$$

When plugged back into the action, it gives rise to the following effective four-point contact interactions (setting $m^2 = 1$ for brevity):

$$\mathcal{L}_{F^4} = \frac{1}{2}[F^4] - \frac{1}{8}[F^2]^2. \tag{6.5}$$

Note that b_2 drops out at this order due to the tracelessness of the lowest-order solution. Moving to the next order, the solution reads

$$h^{(2)} = -2[F^2]F^2 + f_1[F^4]\eta + f_2[F^2]^2\eta, \tag{6.6}$$

---

[2]Theories of massive gravity typically suffer from ghosts, for which it is inconsistent to neglect the Einstein-Hilbert term. We will return to this issue in section 6.4.

for functions $f_{1,2}$ of the parameters (b_2, c_1, c_2) whose specific forms are not relevant for our purposes. We have rewritten an F^4 term using the following recursive matrix identity:

$$X^n = \frac{1}{2}X^{n-1}[X] + \frac{1}{4}X^{n-2}[X^2] - \frac{1}{8}X^{n-2}[X]^2, \ n \geq 2 \,,$$

(6.7)

for $X = F^2$. Moreover, it is also valid for $X = h$ due to the fact that h is solved for in terms of F^2. Upon plugging the solution back into the action, we obtain the six-point contact interactions:

$$\mathcal{L}_{F^6} = \frac{1}{2}[F^4][F^2] - \frac{1}{8}[F^2]^3 \,.$$

(6.8)

Remarkably, all MG coefficients again drop out at this order; ultimately this can again be traced back to the tracelessness of the lowest-order solution $h^{(1)}$, such that the new coefficients only appear in the trace part of $h^{(2)}$ - which drops out at the F^6 order.

The parametric freedom in the MG potential therefore only comes in at octic order: the coefficients of the three terms $[F^4]^2$, $[F^4][F^2]^2$ and $[F^2]^4$ will depend on the first five parameters $(b_2, c_1, c_2, d_1, d_3)$. Moreover, the ten-point interactions are fixed in terms of the eight-point interactions and hence depend on the same combination of these five parameters. Once we move to the twelve- and fourteen-point interactions, four new combinations of nine additional parameters of the graviton potential appear.

We therefore find that massive gravity leads to a very specific non-linear modification of Maxwell's free theory: the massive gravity potential induces a new interaction at every order in F^4, with corresponding terms at higher orders in F^2. As we will see below, these are precisely those theories of non-linear electrodynamics that are invariant under electromagnetic duality.

6.3 Duality-invariant non-linear electrodynamics

Maxwell's equations for a free photon are famously invariant under transformations that rotate the electric and magnetic fields into each other:

$$F_{\mu\nu} \rightarrow \cos\alpha \, F_{\mu\nu} + \sin\alpha \, \tilde{F}_{\mu\nu} \,.$$

(6.9)

These transformations are often referred to as 'duality' rotations between the field strength and its Hodge dual $\tilde{F}_{\mu\nu} \equiv \frac{1}{2}\varepsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}$. Given the central role of symmetries

in high energy physics, a natural question is whether there are non-linear extensions of Maxwell theory that preserve this duality invariance.

The most general theory of this type was constructed by Gibbons and Rasheed[3] [28] and is most conveniently written in terms of two invariants

$$p_\pm = \frac{1}{8}\left([F^2] \pm \sqrt{4[F^4] - [F^2]^2}\right).$$

(6.10)

Duality invariance imposes the following constraint on the Lagrangian:

$$\mathcal{L}_-\mathcal{L}_+ = 1,$$

(6.11)

where $\mathcal{L}_\pm$ denotes the derivative of $\mathcal{L}$ with respect to $p_\pm$. The most general solution to this requirement was subsequently given by Gaillard and Zumino [29] and involves the freedom to specify an arbitrary function $v(s)$. Its argument is solved for in terms of $p_\pm$ via

$$p_+ = v'^2(p_- - s),$$

(6.12)

after which the Lagrangian takes the form

$$\mathcal{L} = \frac{2p_+}{v'} + v.$$

(6.13)

There are two well-known closed-form solutions to (6.12). As a simple generalisation of Maxwell ($v = s$), one can take $v = s \exp(c)$ for some constant c. This leads to the so-called ModMax theory,

$$\mathcal{L} = \frac{1}{4}\cosh(c)[F^2] + \frac{1}{4}\sinh(c)\sqrt{4[F^4] - [F^2]^2}$$

(6.14)

with the special property of being the only modification that, on top of duality invariance, also preserves conformal symmetry [30]. This Lagrangian is not analytic when expanded around an $F = 0$ vacuum, however, and hence can only be used for non-trivial electromagnetic backgrounds. To ensure analyticity around $F = 0$, the Lagrangian has to be even under the interchange of $p_\pm$; we will restrict to this class from now onwards. The second closed form solution, with analyticity, is Born-Infeld theory [31], to which we will turn in the next section. Further generalisations were found in [32].

---

[3]A somewhat similar construction was already presented in [26]. Further details on the history of this theory can be found in [27].

In order to connect to massive gravity, one should first note that the above system of equations (6.12), (6.13) allows for an interpretation of s as an auxiliary field: at the solution for s, the Lagrangian can be written as

$$\mathcal{L} = \frac{1}{v'}p_+ + v'p_- - V_s, \quad V_s = sv' - v. \tag{6.15}$$

Provided $v'' \neq 0$, the constraint (6.12) is then actually equivalent to the field equation for s. In this new formulation, the duality requirement (6.11) is moreover manifestly satisfied. It thus follows that the construction of the most general non-linear duality-invariant electrodynamics corresponds to integrating out an auxiliary field[4] [32, 36]; however, what remains to be explained are the very specific couplings $v'^{\pm 1}$ of s to the two components $p_\pm$ of the Maxwell field.

Remarkably, these follow exactly from integrating out massive gravity, and thus are determined by diffeomorphism invariance and minimal coupling. To show this, we will proceed in three steps. First observe that solving the metric, g, in terms of the field strength, F^2, and a background metric η via the constraint equation (6.3) implies that the matrices corresponding to g and F^2 must commute. In other words, the two are simultaneously diagonalizable and, as a consequence, the eigenvalues of g come in two pairs with distinct values, i.e. $\lambda_{1,2}$. This confirms the intuition one obtains from the perturbative approach of the previous section, where the metric is derived in terms of a polynomial in the field strength; its most general, non-perturbative expression is

$$g = c_1(g, F)\eta + c_2(g, F)F^2, \tag{6.16}$$

where $c_{1,2}$ are scalar functions of the metric and field-strength traces. We can fix these in terms of traces as

$$g = \frac{1}{4}[g]\eta + \sqrt{\frac{[g]^2 - 4[g^2]}{[F^2]^2 - 4[F^4]}} \left(F^2 - \frac{1}{4}[F^2]\eta \right), \tag{6.17}$$

to ensure that the metric obeys the correct relation to the eigenvalues $[g^i] = 2(\lambda_1^i + \lambda_2^i)$.

Secondly, since the Maxwell term in the Lagrangian is conformally invariant, it can only depend on the ratio of these two eigenvalues. The specific expression follows from plugging (6.17) into (6.1) and turns out to be particularly simple:

$$\mathcal{L}_{gF^2} = c_+ p_+ + c_- p_-, \quad c_\pm = (\lambda_1/\lambda_2)^{\pm 1}. \tag{6.18}$$

---

[4]This approach can be extended to supersymmetric systems, see e.g. [33–35].

Note the close similarity to the auxiliary field formulation (6.15); however, we have now *derived*, instead of postulated, the specific couplings to $p_\pm$ from massive gravity.

Finally, in contrast to the Maxwell part, the graviton potential in equation (6.1) is not conformally invariant and therefore depends more generally on the two eigenvalues rather than their ratio only. A convenient approach is to separate off a conformal factor, $g = \rho^2 \tilde{g}$. This factor contributes via the MG potential to the trace (taken with the inverse metric) of the graviton field equation (6.3). Solving for it via this trace equation (or, alternatively, extremising with respect to the conformal factor itself) yields a 'reduced' MG potential with conformal invariance, which only depends on the ratio of eigenvalues (of either g or $\tilde{g}$). Moreover, its dependence has to be invariant under relabelling of the eigenvalues. The simplest quantity with this property is:

$$f^2 \equiv \frac{\lambda_1}{\lambda_2} + \frac{\lambda_2}{\lambda_1} - 2 \,. \tag{6.19}$$

Crucially, this is small if the auxiliary metric is close to flat-space, with the first contributions coming in at quadratic order: $f^2 = [h^2] - \frac{1}{4}[h]^2 + \mathcal{O}(h^3)$; note that odd powers of f are not analytic around $h = 0$. Without loss of generality, the most general EFT can thus be generated by a reduced MG potential,

$$V_f = \sum_n a_n f^{2n} \,, \tag{6.20}$$

where the coefficients, a_i, are generated by the original MG potential in the process of integrating out ρ. Expressing (6.18) in the same combination yields

$$c_\pm = 1 + \frac{1}{2} f^2 \pm \frac{1}{2} \sqrt{f^2(4 + f^2)} \,. \tag{6.21}$$

which preserves duality invariance according to (6.11), since $c_- c_+ = 1$.

At this point it is simple to identify the mapping between the formulations with auxiliary fields s and f^2. For a given $v(s)$, one can relate the two auxiliary fields by $c_- = v'$, which can be plugged into the potential to determine to which V_f this corresponds. Conversely, for a given V_f, the same two equations allow one to solve for v' in terms of v and s, which can then be solved. There is thus a one-to-one mapping between the auxiliary field formulations in terms of $v(s)$ and in terms of V_f. We will provide an explicit example of this in the next section.

6.4   dRGT & Born-Infeld

So far we have considered the most general graviton potential in equation (6.1). However, MG theories typically suffer from negative energy states known as *ghosts*, see e.g. Ref. [37]. In order for our theories to be physical, we therefore have to restrict ourselves to ghost-free theories, requiring a very specific tuning of the coefficients of the general MG potential in equation (6.1), known as dRGT gravity [4, 5].

At the linearised level this amounts to the Fierz-Pauli tuning $b_2 = -1$ [38], while its non-linear extension is defined in terms of a new tensor quantity [5]:

$$\mathcal{K}^\mu{}_\nu = \delta^\mu{}_\nu - \sqrt{\delta^\mu{}_\nu - g^{\mu\alpha}h_{\alpha\nu}}\,, \tag{6.22}$$

or, in matrix notation, $\mathcal{K} = \mathcal{I} - \sqrt{g^{-1}\eta}$. MG in 4D is then ghost-free for a two parameter family of potentials

$$\mathcal{L}_{\Lambda_3} = -\frac{1}{8}m^2\sqrt{-g}\sum_{n=2}^{4}\alpha_n\mathcal{L}_{\mathcal{K}}^{(n)}\,,$$
$$\mathcal{L}_{\mathcal{K}}^{(2)} = [\mathcal{K}]^2 - [\mathcal{K}^2]\,,$$
$$\mathcal{L}_{\mathcal{K}}^{(3)} = [\mathcal{K}]^3 - 3[\mathcal{K}][\mathcal{K}^2] + 2[\mathcal{K}^3]\,,$$
$$\mathcal{L}_{\mathcal{K}}^{(4)} = [\mathcal{K}]^4 - 6[\mathcal{K}^2][\mathcal{K}]^2 + 8[\mathcal{K}^3][\mathcal{K}] + 3[\mathcal{K}^2]^2 - 6[\mathcal{K}^4]\,, \tag{6.23}$$

where $\alpha_i = (-4, 2^3c_3, 2^4d_5)$ and we will explicitly include the graviton mass m in this section. Another important consequence of this tuning is the fact that the UV-cutoff of the theory is raised from $\Lambda_5 = (M_\mathrm{P}m^4)^{1/5}$ to $\Lambda_3 = (M_\mathrm{P}m^2)^{1/3}$. This theory is therefore also referred to as Λ_3-theory.

We will demonstrate how dRGT gravity is connected to a very specific non-linear completion due to Born and Infeld [31], which has the remarkable property that it can be formulated geometrically:

$$\mathcal{L}_\mathrm{BI} = -\frac{m^2}{4}\left(\left(-\det(\eta - \frac{4}{m^2}F^2)\right)^{1/4} - 1\right)\,. \tag{6.24}$$

It arises in a variety of contexts, including string theory (as the world-volume theory of D-branes [13]), the double copy (allowing for a CHY formulation of its amplitudes [14]) and soft limits (under a multi-chiral soft limit [15]). This prompts the question whether it follows from the ghost-free dRGT theory and, if so, what its graviton potential is.

To this end, we will first derive the reduced graviton potential corresponding to Born-Infeld theory. Following the example of Ref. [29], the auxiliary field formulation of Born-Infeld in terms of s is[5]

$$v(s) = -\frac{m^2}{4}\left(\sqrt{1 - \frac{8}{m^2}s} - 1\right).$$
(6.25)

Following the general procedure outlined in the previous section, this can be mapped onto the following auxiliary field formulation in terms of f^2:

$$V_f = \frac{m^2}{8}\frac{(v'-1)^2}{v'} = \frac{m^2}{8}f^2,$$
(6.26)

where in the last equality we have made use of the relation (6.21). Hence, we conclude that the simplest possible reduced MG potential, that is only linear in f^2, leads to Born-Infeld.

It remains to see whether this fits within dRGT. Following the perturbative approach of section 6.2 and matching order by order, we find that Born-Infeld theory requires the dRGT parameters $c_3 = \frac{1}{6}$ and $d_5 = 0$. For these values, the full potential can be written as

$$\frac{1}{m^2}\mathcal{L}_{\Lambda_3} = 1 + \sqrt{-g}\Big([\mathcal{K}] - \det(\mathcal{K}) - 1\Big),$$
(6.27)

where we have used $\det(\mathcal{I} - \mathcal{K}) = \det(g)^{-1/2}$. To integrate out the graviton, we follow the procedure described in section 6.3. First, we rewrite the above Lagrangian in terms of the metric,

$$\frac{1}{m^2}\mathcal{L}_{\Lambda_3} = 1 + \sqrt{-g}\Big(2 + [g^{-1/2}]^2(-\frac{1}{2} + \frac{1}{8}[g^{-1/2}] - \frac{1}{64}[g^{-1/2}]^2) + [g^{-1}](\frac{1}{2} - \frac{1}{4}[g^{-1/2}]$$
$$+ \frac{1}{16}[g^{-1/2}]^2 - \frac{1}{16}[g^{-1}])\Big),$$
(6.28)

where $g^{-1/2} \equiv \sqrt{g^{-1}\eta}$. We then pull out a conformal factor from the metric, $g = \rho^2\tilde{g}$ and solve for it using the trace of (6.3), or alternatively simply vary with respect to ρ itself. For the specific potential of (6.28) this yields the particularly simple solution $\rho = [\tilde{g}^{-1/2}]$. It follows naturally that the resulting reduced MG potential is conformally invariant which allows us to work with the original metric g again. After this substitution, the full Lagrangian can be compactly written as

$$\mathcal{L} = \frac{1}{4}\sqrt{-g}\langle F^2\rangle - \frac{m^2}{16}\sqrt{-g}\langle\eta^2\rangle + \frac{m^2}{4},$$
(6.29)

---

[5]Notice a sign difference compared to Ref. [29], which is due to their p, q corresponding to $-p_-, -p_+$ in our notation.

where $\langle \eta^2 \rangle = [g^{-2}]$ in our notation. Upon varying and solving (6.29) for g, with the simple solution

$$g = \eta \sqrt{ \mathbb{I} - \frac{4}{m^2} F \cdot \eta \cdot F } \, , \qquad (6.30)$$

this results precisely in Born-Infeld theory with the coupling scale set by the mass of the graviton as given in equation (6.24).

We thus find that Born-Infeld follows from integrating out the massive graviton in a specific dRGT theory. Moreover, the relevant part of the MG potential in (6.29) indeed takes the simple form (6.26) predicted by the auxiliary field prescription. From this perspective, Born-Infeld corresponds to the leading non-linear completion of Maxwell theory, with the simplest possible conformally invariant potential f^2. Higher-order terms of f^2 in the reduced graviton potential encode the other parameter choices of dRGT.

As a corollary remark, we note that for the choice of parameters $c_3 = \frac{1}{6}$ and $d_5 = 0$, the helicity-0 component of the theory in the decoupling limit, generally given by the cubic, quartic and quintic Galileon terms, now reduce to the quartic special Galileon (see e.g. equation (9.27) in Ref. [39]).

6.5 Conclusions

In this chapter we have studied the low energy effective action of Maxwell theory coupled to massive gravity by integrating out the massive graviton. The invariance of Maxwell theory under duality rotations is retained in this scenario, as gravity is unaffected by such transformations. Hence, we have found a reformulation of the previous classification of duality invariant non-linear electrodynamics [28, 29] in terms of an auxiliary graviton field[6]. The conformal invariance of Maxwell plays an important role in this approach, and allows the MG potential to be truncated to a conformal section of it. Remarkably, it turns out that one can generate all non-linear, analytic completions of Maxwell[7] in this manner. Although an auxiliary field formulation of non-linear electrodynamics is not new (e.g. [32, 36]), the specific couplings that are needed now follow from minimal coupling and thus diffeomorphism invariance, instead of being included *ad hoc*. Therefore, our approach naturally preserves

---

[6]Note that auxiliary metrics have also been introduced in the construction of self-dual p-form actions [40, 41]; it would be interesting to investigate possible connections.

[7]Perhaps the same applies to ModMax, such that its non-linear Born-Infeld completion [42] can be derived by coupling it to a massive graviton. We leave this for future work.

principles such as locality and, moreover, has an obvious generalization to include derivative interactions.

Restricting ourselves to the subset of ghost-free massive gravity theories known as dRGT gravity, we identified a particular theory whose conformal MG potential takes a simple quadratic form in terms of the relabeling-invariant combination (6.19) of the graviton's eigenvalues. Integrating out the remaining graviton components leads to the geometric Born-Infeld action to all orders. Given the latter's special properties in terms of amplitudes, it would be interesting to investigate what the current perspective implies. For instance, we have restricted ourselves to the lowest order in the expansion of scattering energy over graviton mass; keeping subleading terms would correspond to higher-derivative terms, e.g. of the form $\partial^4 F^4$. Can these be formulated in terms of curvature invariants of the metric $\eta + F$? Similarly, do these higher-derivative corrections fit in the CHY formalism [14, 43]?

Furthermore, one can wonder whether there is something special about non-linear completions obtained from dRGT massive gravity, since these follow from integrating out a well defined physical graviton field, compared to theories obtained from non-dRGT massive gravity, where one naively integrates out an auxiliary field that propagates ghosts. For example, it would be interesting to see whether the additional dRGT parameters (beyond the specific point with the MG potential (6.28)) fit in the double copy framework, and moreover whether there is a connection to positivity bounds on these theories [11, 12] as well.

On a different note, given the crucial role of conformal invariance in our approach, it is natural to ask whether a similar construction can be applied to a scalar field in 2D. In fact, this case is even more special: massive gravity does not propagate any physical degrees of freedom in two dimensions and is therefore a genuine auxiliary field. Moreover, the dRGT graviton potential has no free parameters in 2D. Indeed it has been found that it maps uniquely to the (multi-field) Dirac-Born-Infeld (DBI) theory by eliminating the non-dynamical degrees of freedom [19]. One can view the above as the massive extension of the usual transition from the Polyakov to the Nambu-Goto form of the string world-sheet action. In the massless case, the auxiliary world-sheet metric is only determined up to an overall factor, due to the conformal symmetry. The introduction of the dRGT potential in 2D breaks this symmetry and can be seen as a gauge fixing term. Integrating out the world-sheet metric now also fixes its overall factor, but this does not change the resulting expressions: the overall factor drops out of the massless part, and moreover the dRGT potential exactly vanishes for this choice (as it acts as a gauge fixing term for the conformal symmetry).

DBI is special: it describes the longitudinal modes of D-branes and non-linearly realises a higher-dimensional version of Poincaré; as a consequence, its amplitudes display a generalised Adler zero in their soft limit [44]. It would be interesting to see whether one can also generate the unique Volkov-Akulov fermionic interactions [45] with these properties in this way.

Given the connection between integrating out dRGT gravity and $T\bar{T}$-deformations in 2D [19], it would be interesting to investigate the possible relation of our approach to $T\bar{T}$-deformations in 4D, see e.g. [20, 21] for Maxwell and [22–24] for ModMax.

In contrast to the special cases discussed here, one loses conformal invariance in generic dimensions. One can easily check (e.g. by a perturbative approach at low orders) that naively integrating out a massive graviton in such cases does not lead to the specific turning of Born-Infeld for a vector field (or DBI for a scalar field), again underlining the importance of conformal invariance. In order to move into other dimensions and theories, it might be necessary to include dilaton interactions, on top of gravity; this combination is natural from e.g. the common sector of string theory, as well as the double copy [16, 43]. However, these set-ups are normally massless; our approach would require an extension of massive gravity to also include a dilaton. We leave these considerations for future work.

6

Bibliography

[1] E. Fermi, "An attempt of a theory of beta radiation. 1.," *Z. Phys.* **88**, pp. 161–177, 1934.

[2] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons," *Phys. Rev. Lett.* **13**, pp. 508–509, 1964.

[3] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons," *Phys. Rev. Lett.* **13**, pp. 321–323, 1964.

[4] C. de Rham and G. Gabadadze, "Generalization of the Fierz-Pauli Action," *Phys. Rev. D* **82**, p. 044020, 2010.

[5] C. de Rham, G. Gabadadze, and A. J. Tolley, "Resummation of Massive Gravity," *Phys. Rev. Lett.* **106**, p. 231101, 2011.

[6] C. Cheung and G. N. Remmen, "Positive Signs in Massive Gravity," *JHEP* **04**, p. 002, 2016.

[7] B. Bellazzini, G. Isabella, S. Ricossa, and F. Riva, "Massive Gravity is not Positive," 4 2023.

[8] N. Beisert, H. Elvang, D. Z. Freedman, M. Kiermaier, A. Morales, and S. Stieberger, "E7(7) constraints on counterterms in N=8 supergravity," *Phys. Lett. B* **694**, pp. 265–271, 2011.

[9] G. Bossard, C. Hillmann, and H. Nicolai, "E7(7) symmetry in perturbatively quantised N=8 supergravity," *JHEP* **12**, p. 052, 2010.

[10] W. Heisenberg and H. Euler, "Consequences of Dirac's theory of positrons," *Z. Phys.* **98**(11-12), pp. 714–732, 1936.

[11] J. Henriksson, B. McPeak, F. Russo, and A. Vichi, "Rigorous bounds on light-by-light scattering," *JHEP* **06**, p. 158, 2022.

[12] M. Carrillo González, C. de Rham, S. Jaitly, V. Pozsgay, and A. Tokareva, "Positivity-causality competition: a road to ultimate EFT consistency constraints," 7 2023.

[13] J. Polchinski, "Dirichlet Branes and Ramond-Ramond charges," *Phys. Rev. Lett.* **75**, pp. 4724–4727, 1995.

[14] F. Cachazo, S. He, and E. Y. Yuan, "Scattering Equations and Matrices: From Einstein To Yang-Mills, DBI and NLSM," *JHEP* **07**, p. 149, 2015.

[15] C. Cheung, K. Kampf, J. Novotny, C.-H. Shen, J. Trnka, and C. Wen, "Vector Effective Field Theories from Soft Limits," *Phys. Rev. Lett.* **120**(26), p. 261602, 2018.

[16] Z. Bern, J. J. M. Carrasco, and H. Johansson, "Perturbative Quantum Gravity as a Double Copy of Gauge Theory," *Phys. Rev. Lett.* **105**, p. 061602, 2010.

[17] D. de Neeling, D. Roest, and S. Veldmeijer, "Flavour-kinematics duality for Goldstone modes," *JHEP* **10**, p. 066, 2022.

[18] C. Cheung, C.-H. Shen, and C. Wen, "Unifying Relations for Scattering Amplitudes," *JHEP* **02**, p. 095, 2018.

[19] A. J. Tolley, "$T\bar{T}$ deformations, massive gravity and non-critical strings," *JHEP* **06**, p. 050, 2020.

[20] A. Cavaglià, S. Negro, I. M. Szécsényi, and R. Tateo, "$T\bar{T}$-deformed 2D Quantum Field Theories," *JHEP* **10**, p. 112, 2016.

[21] R. Conti, L. Iannella, S. Negro, and R. Tateo, "Generalised Born-Infeld models, Lax operators and the $T\bar{T}$ perturbation," *JHEP* **11**, p. 007, 2018.

[22] H. Babaei-Aghbolagh, K. B. Velni, D. M. Yekta, and H. Mohammadzadeh, "Emergence of non-linear electrodynamic theories from $T\bar{T}$-like deformations," *Phys. Lett. B* **829**, p. 137079, 2022.

[23] C. Ferko, L. Smith, and G. Tartaglino-Mazzucchelli, "On Current-Squared Flows and ModMax Theories," *SciPost Phys.* **13**(2), p. 012, 2022.

[24] R. Conti, J. Romano, and R. Tateo, "Metric approach to a $T\bar{T}$-like deformation in arbitrary dimensions," *JHEP* **09**, p. 085, 2022.

[25] K. Hinterbichler, "Theoretical Aspects of Massive Gravity," *Rev. Mod. Phys.* **84**, pp. 671–710, 2012.

[26] I. Bialynicki-Birula, "Nonlinear Electrodynamics: Variations on a theme by Born and Infeld," in *B. Jancewicz and J. Lukierski eds/, Quantum Theory of Particles and Fields, World Scientific*, pp. 31–48, 1983.

[27] D. P. Sorokin, "Introductory Notes on Non-linear Electrodynamics and its Applications," *Fortsch. Phys.* **70**(7-8), p. 2200092, 2022.

[28] G. W. Gibbons and D. A. Rasheed, "Electric - magnetic duality rotations in nonlinear electrodynamics," *Nucl. Phys. B* **454**, pp. 185–206, 1995.

[29] M. K. Gaillard and B. Zumino, "Nonlinear electromagnetic selfduality and Legendre transformations," in *A Newton Institute Euroconference on Duality and Supersymmetric Theories*, pp. 33–48, 12 1997.

[30] I. Bandos, K. Lechner, D. Sorokin, and P. K. Townsend, "A non-linear duality-invariant conformal extension of Maxwell's equations," *Phys. Rev. D* **102**, p. 121703, 2020.

[31] M. Born and L. Infeld, "Foundations of the new field theory," *Proc. Roy. Soc. Lond. A* **144**(852), pp. 425–451, 1934.

6

[32] M. Hatsuda, K. Kamimura, and S. Sekiya, "Electric magnetic duality invariant La-grangians," *Nucl. Phys. B* **561**, pp. 341–353, 1999.

[33] S. M. Kuzenko and S. Theisen, "Supersymmetric duality rotations," *JHEP* **03**, p. 034, 2000.

[34] E. Ivanov, O. Lechtenfeld, and B. Zupnik, "Auxiliary superfields in N=1 supersymmetric self-dual electrodynamics," *JHEP* **05**, p. 133, 2013.

[35] J. J. M. Carrasco, R. Kallosh, and R. Roiban, "Covariant procedures for perturbative non-linear deformation of duality-invariant theories," *Phys. Rev. D* **85**, p. 025007, 2012.

[36] E. A. Ivanov and B. M. Zupnik, "New representation for Lagrangians of selfdual nonlin-ear electrodynamics," in *4th International Workshop on Supersymmetry and Quantum Sym-metries: 16th Max Born Symposium*, pp. 235–250, 2002.

[37] D. G. Boulware and S. Deser, "Can gravitation have a finite range?," *Phys. Rev. D* **6**, pp. 3368–3382, 1972.

[38] M. Fierz and W. Pauli, "On relativistic wave equations for particles of arbitrary spin in an electromagnetic field," *Proc. Roy. Soc. Lond. A* **173**, pp. 211–232, 1939.

[39] K. Hinterbichler and A. Joyce, "Hidden symmetry of the Galileon," *Phys. Rev. D* **92**(2), p. 023503, 2015.

[40] A. Sen, "Self-dual forms: Action, Hamiltonian and Compactification," *J. Phys. A* **53**(8), p. 084002, 2020.

[41] C. M. Hull, "Covariant Action for Self-Dual p-Form Gauge Fields in General Space-times," 7 2023.

[42] I. Bandos, K. Lechner, D. Sorokin, and P. K. Townsend, "On p-form gauge theories and their conformal limits," *JHEP* **03**, p. 022, 2021.

[43] F. Cachazo, S. He, and E. Y. Yuan, "Scattering of Massless Particles in Arbitrary Dimen-sions," *Phys. Rev. Lett.* **113**(17), p. 171601, 2014.

[44] C. Cheung, K. Kampf, J. Novotny, and J. Trnka, "Effective Field Theories from Soft Limits of Scattering Amplitudes," *Phys. Rev. Lett.* **114**(22), p. 221602, 2015.

[45] D. V. Volkov and V. P. Akulov, "Is the Neutrino a Goldstone Particle?," *Phys. Lett. B* **46**, pp. 109–110, 1973.

Chapter 7

Discussion and Outlook

Cosmology is about to enter an era of unprecedented amounts of observational data, using a variety of probes, such as the relic Cosmic Microwave Background, the distribution of galaxies across the low-redshift universe, and the 21-cm neutral hydrogen line across the universe's entire history. This wealth of data is poised to lead to transformative insights into the physics governing the universe at the largest scales, thereby hopefully resolving some of the open questions and tensions in cosmology, including the mysterious nature of Dark Matter and Dark Energy. We also hope to gain insight into the fundamental physics on the very smallest length scales and time scales too, possibly shedding light on already existing hints of physics beyond the standard model, such as massive neutrinos [1]. Additionally, we will be able to perform the most stringent test of the inflationary paradigm so far, including a possible detection of CMB B-modes.

Among these tests of inflation, primordial non-Gaussianity constitutes one of the most promising and exciting windows into the early universe. Its ability to discern physical mechanisms that drive the exponential expansion of inflation, as well as its sensitivity to the particle content of this epoch, make it a unique probe of fundamental physics at an unprecedented energy scale. So far, extensive efforts to map the CMB to ever-increasing precision, have not resulted in conclusive evidence for the existence of these non-Gaussian features, but have led to tight constraints as to what their magnitude can be, and we will have another shot at a detection with the upcoming next generation of surveys. Soon though, the lack of anisotropies on the very smallest scales of the CMB will force us to look beyond this otherwise rich source of data.

The large-scale structure of the universe spans a vast three-dimensional volume covering different epochs from the present up to the CMB surface, containing orders of magnitude more information than the two-dimensional CMB image. The most recent of these epochs we will be able to probe by studying the spatial distribution of galaxies across the universe as a tracer of the matter distribution. Although this will provide us with a vast amount of data, the non-linear evolution of the mat-

ter distribution and the relatively unknown process of structure formation severely complicate the extraction of valuable cosmological information as compared to the analysis of the CMB, which is both highly Gaussian and linearly related to the initial conditions seeded by inflation. Since only the largest scales of our present-day universe can currently be modeled with sufficient accuracy, state-of-the-art cosmological analysis of galaxy data uses only a subset of the total information present in the data. The tools to perform such analyses, both analytical or based on numerical simulations, are under active development and it remains yet to be seen how much of the data can be used to constrain further our cosmological models, including primordial non-Gaussianity.

The problem of non-linearity is naturally less severe at earlier times, i.e. higher redshifts, before the complicated physics of forming bound structures took place. For these earlier epochs, we hope to map the distribution of neutral hydrogen through its 21-cm spectral line as a tracer of the underlying matter distribution. In particular, the earliest of these epochs, the Dark Ages, stands out both since it is highly linear and virtually free of any unknown astrophysics, in principle making it the cleanest probe for cosmology after the CMB. This therefore promises the Dark Ages' 21-cm signal to also be an excellent probe of primordial non-Gaussianity. In Chapter 2, we have investigated this claim quantitatively. After having established the most accurate analytical model of the 21-cm signal from the Dark Ages, we use it to forecast constraints on primordial non-Gaussianity, for the first time including the primordial trispectrum, while carefully accounting for the confusion of the primordial signal with intrinsic non-Gaussianity of the tracer through marginalization. We confirm the probe's unprecedented sensitivity to primordial non-Gaussianity and its potential to provide a new window into fundamental physics at the highest energy scales, paving the way for cosmological collider physics.

As already outlined in Chapter 1, the challenge that a measurement of the 21-cm signal from the Dark Ages poses, is daunting. Strong foreground contamination (including Earth's ionosphere) severely complicates the detection of the minute 21-cm signal, let alone its even smaller fluctuations that allow us to trace the matter distribution and learn about the initial conditions of the universe. Nevertheless, there has been serious effort by the scientific community to propose future space- or lunar-based missions with a focus on exploring the cosmic Dark Ages [2–6], some of which are being actively considered by international space agencies [7, 8]. Given the extraordinary capacity and perseverance of scientists, the author is hopeful that the futuristic science described in Chapter 2 will eventually become a reality.

Meanwhile, even if we can model highly linear tracers such as the 21-cm sig-

nal from the Dark Ages accurately out to very small scales, this does not set us free from the consequences of non-linearity. Fourier modes of different wavelengths are inevitably coupled by the non-linear evolution that they undergo. Hence, also the summary statistics of these modes that we hope to analyze (the data), are correlated. This non-Gaussian covariance results in a reduced amount of unique information contained in the data and needs to be properly accounted for both during inference, as well as when making experimental forecasts. Failure to do so will result in overconfident constraints and overly optimistic expectations of future experimental endeavors. However, many of the existing forecasts on primordial non-Gaussianity using high-redshift tracers tend to neglect its effects, assuming that it is small.

In Chapter 3, we investigated the validity of this assumption and we study the impact of non-Gaussian covariance on forecast constraints on primordial non-Gaussianity. We confirm the recent finding that non-Gaussian contributions to the bispectrum covariance are largest for highly squeezed triangles, thus significantly impacting the constraining power of the bispectrum for local (squeezed) primordial non-Gaussianity [1]. Additionally, we show for the first time that as we go to higher redshift, where we expect to be able to access smaller scales and thus increasingly squeezed triangles, the impact on local primordial non-Gaussianity increases, and causes the constraining power of the bispectrum to saturate. This goes against the naïve expectation that for more linear tracers, non-Gaussian covariance is by definition a smaller effect. To demonstrate how this affects observational constraints, we revised the original forecast of the PUMA 21-cm line-intensity-mapping experiment, which neglects non-Gaussian covariance. Most importantly, we find that accounting for this additional covariance increases forecast error bars on local primordial non-Gaussianity using the hydrogen bispectrum, by a factor of 2 to 5 depending on the redshift considered. The results of the chapter reveal a previously overlooked complication in using high redshift ($z > 2$) tracers to constrain local primordial non-Gaussianity, and hence careful interpretation of forecasts such as those presented in Chapter 2 is warranted.

A recent work that the author of this thesis has contributed to, explores an alternative way of using the 21-cm signal from the Dark Ages [10]. Since the scales accessed in the CMB are of a much larger physical size than those accessed with the 21-cm signal from the Dark Ages, one can combine the two tracers to probe highly squeezed triangle configurations, making it an attractive setup for measuring local primordial non-Gaussianity. Furthermore, at the time of CMB emission, and on

---

[1]It has furthermore been shown that even if one can model the matter bispectrum to far within the non-linear regime using consistency relations, non-Gaussian covariance of the bispectrum quickly satures constraints on local primordial non-Gaussianity beyond the non-linear scale [9].

scales as large as the lowest CMB multipoles, the size of matter fluctuations is small. This reduces the strength of non-Gaussian covariance on these squeezed triangles, and it is shown that using cross-correlations of CMB and 21-cm from the Dark Ages can yield exquisite constraints on local primordial non-Gaussianity. This provides additional motivation for future efforts to map the 21-cm signal from the Dark Ages.

Non-Gaussian covariance of summary statistics is a manifestation of the fact that non-linear evolution has moved information from its starting point (e.g. in squeezed triangles) to smaller scales and higher-order spatial correlation functions. Thus capturing all information in the data requires computation of such correlation functions, which can quickly become more expensive. This is in part because for a larger data vector, one requires significantly more simulations to obtain an accurate covariance matrix, and the inclusion of higher-order correlations makes theoretical covariance matrix computation infeasible too. In Chapter 4, we investigated to which extent this problem can be alleviated by reconstructing the linear matter distribution at the field level. The goal of the reconstruction is to move information back into the lower-order correlation functions, whose information content is then analyzed. We find that a reconstructive approach using convolutional neural networks can indeed improve constraints on cosmological parameters. In particular, constraints on primordial non-Gaussianity are greatly improved, as the information content of mainly the bispectrum is significantly increased. Our results show that cosmological analysis of data from the late-time non-linear universe can be enhanced by going beyond the direct use of summary statistics.

Instead, our approach can be regarded as a form of *field-level inference*, as our network employs the unsummarized data of the matter density field (thus limited only by pixel resolution). Over recent years this type of inference has gained attention, which is in part due to rapid developments in the field of artificial intelligence (AI), that allow the handling of such high-dimensional data without the need for summary statistics. For example, graph neural networks enable extraction of cosmological parameters directly from simulated galaxy (or halo) catalogs [11, 12]. Another approach, that is particularly interesting in the context of primordial non-Gaussianity, is that of simultaneously inferring (i.e. reconstructing) the initial conditions and other cosmological parameters from non-linear data (e.g. the late-time matter distribution) using a (Bayesian) posterior probability distribution and a forward model that simulates this non-linear data given some initial conditions and cosmological parameters. With traditional Bayesian methods, such as Markov Chain Monte Carlo, this high-dimensional problem is intractable, due to the sheer size of the parameter space that needs to be explored. However, if we were to have access to gradients (i.e. derivatives) of the posterior, it would allow for both efficient optimiza-

tion of the most likely initial conditions and cosmological parameters (maximizing the posterior), as well as the use of gradient-based Bayesian sampling techniques, such as Hybrid (Hamiltonian) Monte Carlo, that can handle such high-dimensional problems. In the case of an analytical (but approximate) forward model for the non-linear evolution, one can obtain this gradient analytically and use it to efficiently sample initial conditions [13]. Such a setup can then also be used to simultaneously infer cosmological parameters [14], including the amplitude of local primordial non-Gaussianity [15]. For more sophisticated forward models, like approximate N-body simulators (e.g. particle-mesh), this gradient cannot be obtained analytically in a straightforward way. Nevertheless, recent advances in AI have led to the development of software that can perform exact differentiation (i.e. not based on finite-difference methods) of such models, and thus the posterior distribution [16, 17]. This again enables Bayesian inference of cosmological initial conditions and cosmological parameters [18]. Given the already significant gains in cosmological parameter constraints achievable with neural network based reconstruction, as described in Chapter 4, it would be interesting to investigate to what extent a Bayesian forward modeling approach can improve upon this. As previous investigations have shown, such an approach can perform optimal reconstruction [16], and thus provides another promising avenue for more optimal extraction of primordial non-Gaussianity from large-scale structure data that deserves further exploration.

In the years to come, the CMB will remain an important source of cosmological information. The fidelity of upcoming CMB experiments such as Simons Observatory and CMB-S4, are poised to determine cosmological parameters with unprecedented accuracy. One complication in analyzing the CMB is that its observed fluctuations are lensed by the late-time matter distribution. This affects our ability to infer the existence of primordial tensor fluctuations and primordial non-Gaussianity, among others. The next generation of CMB surveys will reach noise levels at which conventional analysis techniques become suboptimal. To extract the maximum amount of information, more sophisticated analyses are thus required. To this end, in Chapter 5, we developed a score-based generative model for performing Bayesian reconstruction of the CMB lensing potential. Most importantly, our model bypasses the need for analytical expressions of the lensing posterior distribution, allowing us to for the first time perform accurate reconstruction of non-Gaussian potentials. The developed methods pave the way for more optimal analysis of future CMB data, as well as the use of the lensing potential as an unbiased tracer of the total matter density distribution.

The aforementioned developments are promising and, combined with rapid advancements in computational technology, will eventually enable the optimal extrac-

tion of all cosmological information from the observable universe. This will in turn drastically improve our understanding of the cosmos and the fundamental principles that govern it, possibly revealing hints, or even evidence, for new and unexplored physics. The hunt for primordial non-Gaussianity provides a promising avenue for such discoveries, and it is only a matter of time until we unravel how skewed the universe really is.

7

Bibliography

[1] J. Lesgourgues and S. Pastor, "Massive neutrinos and cosmology," *Phys. Rept.* **429**, pp. 307–379, 2006.

[2] S. Furlanetto *et al.*, "Astro 2020 Science White Paper: Fundamental Cosmology in the Dark Ages with 21-cm Line Fluctuations," 3 2019.

[3] J. O. Burns *et al.*, "Dark Cosmology: Investigating Dark Matter \& Exotic Physics in the Dark Ages using the Redshifted 21-cm Global Spectrum," 2 2019.

[4] J. O. Burns, "Transformative science from the lunar farside: observations of the dark ages and exoplanetary systems at low radio frequencies," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **379**, p. 20190564, Nov. 2020.

[5] X. Chen *et al.*, "Discovering the sky at the longest wavelengths with small satellite constellations," 2019.

[6] L. V. E. Koopmans *et al.*, "Peering into the dark (ages) with low-frequency space interferometers," 2019.

[7] J. O. Burns *et al.*, "Nasa probe study report: Farside array for radio science investigations of the dark ages and exoplanets (farside)," 2019.

[8] A. Goel *et al.*, "Probing the Cosmic Dark Ages with the Lunar Crater Radio Telescope," 5 2022.

[9] S. Goldstein, A. Esposito, O. H. E. Philcox, L. Hui, J. C. Hill, R. Scoccimarro, and M. H. Abitbol, "Squeezing fNL out of the matter bispectrum with consistency relations," *Phys. Rev. D* **106**(12), p. 123525, 2022.

[10] G. Orlando, T. Flöss, P. D. Meerburg, and J. Silk, "Local non-Gaussianities from cross-correlations between the CMB and 21-cm," 7 2023.

[11] H. Shao *et al.*, "Robust Field-level Inference of Cosmological Parameters with Dark Matter Halos," *Astrophys. J.* **944**(1), p. 27, 2023.

[12] N. S. M. de Santi *et al.*, "Robust Field-level Likelihood-free Inference with Galaxies," *Astrophys. J.* **952**(1), p. 69, 2023.

[13] J. Jasche and B. D. Wandelt, "Bayesian physical reconstruction of initial conditions from large scale structure surveys," *Mon. Not. Roy. Astron. Soc.* **432**, p. 894, 2013.

[14] N.-M. Nguyen, F. Schmidt, B. Tucci, M. Reinecke, and A. Kostić, "How much information can be extracted from galaxy clustering at the field level?," 3 2024.

[15] A. Andrews, J. Jasche, G. Lavaux, and F. Schmidt, "Bayesian field-level inference of primordial non-Gaussianity using next-generation galaxy surveys," *Mon. Not. Roy. Astron. Soc.* **520**(4), pp. 5746–5763, 2023.

[16] Y. Feng, U. Seljak, and M. Zaldarriaga, "Exploring the posterior surface of the large scale structure reconstruction," *JCAP* **07**, p. 043, 2018.

[17] Y. Li, C. Modi, D. Jamieson, Y. Zhang, L. Lu, Y. Feng, F. Lanusse, and L. Greengard, "Differentiable Cosmological Simulation with the Adjoint Method," *Astrophys. J. Suppl.* **270**(2), p. 36, 2024.

[18] A. E. Bayer, U. Seljak, and C. Modi, "Field-Level Inference with Microcanonical Langevin Monte Carlo," in *40th International Conference on Machine Learning*, 7 2023.

7

Summary

Some of the most fundamental open questions in physics pertain to what happened during the first moments after the Big Bang when the universe came into existence. One of the most well-established theories is that the universe underwent a period of rapidly accelerated (exponential) expansion for a tiny fraction of a second, expanding space by upwards of twenty orders of magnitude. This period goes by the name of **cosmic inflation**. The most compelling hints for such an expansion having occurred come from the Cosmic Microwave Background (CMB) radiation. This relic light, emitted only 300.000 years after the Big Bang, provides us with an early snapshot of the state of the early universe. The inflationary expansion helps to explain the observational fact that two causally disconnected regions in the CMB today (e.g. two polar opposite directions in the sky) have a nearly identical temperature. This apparent *fine-tuning* of the universe can instead be explained if the observable universe started out much smaller, such that all of it was in causal contact early on and thus able to reach a thermal equilibrium. The CMB is not perfectly homogeneous and isotropic though, and we have measured tiny temperature fluctuations of 1 part in 10000, that appear to be Gaussian distributed. Strikingly, cosmic inflation also provides a mechanism for generating these tiny anisotropies that we see in the CMB, and that serve as the seeds of structure formation and can thus be thought of as the initial conditions of the universe, by means of quantum fluctuations.

Inflationary expansion can be physically realized in a myriad of ways, many of which are compatible with our precise observations of the CMB to date. To advance our understanding of the early universe, we thus need to look for additional ways to discriminate different inflationary scenarios. One such way is through what is known as **primordial non-Gaussianity**. Although so far our measurements of the CMB tell us that the initial conditions appear to be Gaussian, models of inflation generically predict some amount of deviation from this. Non-Gaussianity of the

initial condition implies the presence of higher-order correlation functions beyond the two-point correlation that characterizes a Gaussian distribution. Contrary to the two-point correlation (or power spectrum in Fourier space), which only depends on the distance between two points (or wavelength in Fourier space), correlations between more than two points, are captured by polygons. Remarkably, different inflationary mechanisms induce correlations in distinctly shaped polygons, thus providing a direct probe of inflation. Focussing on the lowest order of non-Gaussian correlations, a three-point function (or bispectrum in Fourier space), we are led to study correlations in various shapes of triangles, as visualized in Figure 7.2.
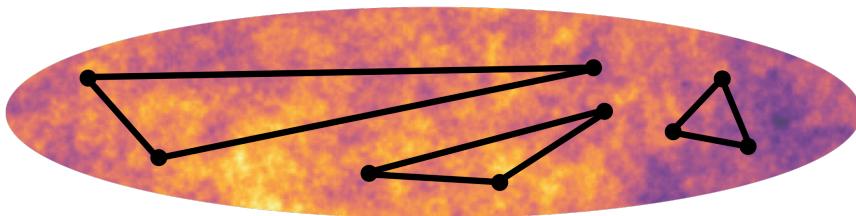

Figure 7.1: *The distribution of primordial density fluctuations, that serve as the initial conditions of our universe, can include higher-order correlation functions, such as three-point correlations. Different inflationary mechanisms induce correlations in distinctly shaped triangles, such as the squeezed, equilateral, and flattened shapes visualized here.*

The simplest theoretical model of cosmic inflation requires the presence of a single scalar field with a suitably shaped potential energy curve, minimally interacting with gravity. The quantum fluctuations of this scalar field translate into the primordial density perturbations at the end of inflation. In this scenario, the amount of primordial non-Gaussianity is predicted to be small. However, direct self-interactions of this scalar field can generate a sizable bispectrum in equilateral triangle configurations. On the contrary, the presence of additional scalar fields contributing to the inflationary expansion is likely to induce a large bispectrum in squeezed triangle configurations. It is thus clear that primordial non-Gaussianity could provide valuable insights into the dynamics of inflation. Besides being a direct probe of the inflationary mechanism, primordial non-Gaussianity is also sensitive to the interactions of the scalar field(s) with additional fields present during (but not contributing to) inflation, the presence of which induces a striking oscillatory feature in the aforementioned squeezed triangles, the amplitude and frequency of which is related to the mass and spin of the spectator field. It is for this reason that inflation can be regarded as a **cosmological particle collider**, the outcome of which has set the initial conditions of the universe, and the energy scale of which is orders of magnitude

larger than that achieved with any terrestrial particle collider experiment. In this way, primordial non-Gaussianity provides a unique and powerful window into the physics at play during the inflationary epoch.

Since we do not have direct access to the initial conditions, we need to infer its properties from the observable universe that these initial conditions gave rise to. As previously mentioned, the CMB provides a pristine tracer of the primordial density fluctuations and has therefore been an indispensable source of information on the early universe. The next generation of CMB surveys aims to provide an even higher resolution view of the early universe, improving constraints on primordial non-Gaussianity in the process. In doing so, these surveys will approach the limit of what the CMB can tell us about primordial non-Gaussianity. Thus we are forced to move our attention toward the distribution of matter throughout the universe, known as **the large-scale structure of the universe**. Although in principle the three-dimensional distribution of matter contains many times more information than the two-dimensional CMB, its relation to the initial conditions becomes more non-linear over time, due to the gravitational collapse of the matter into bound structures (e.g. galaxies). Extracting primordial non-Gaussianity from the matter distribution therefore comes with substantial difficulties. This thesis presents several advancements, challenges and prospects in the hunt for primordial non-Gaussianity using the large-scale structure of the universe.

In Chapter 2 we investigate the use of the 21-centimeter hyperfine transition of the neutral hydrogen gas that permeates the universe during the **Cosmic Dark Ages**, as a probe of primordial non-Gaussianity. We find that this pristine tracer of primordial fluctuations, together with its vast three-dimensional volume, yields unparalleled precision to detect or constrain primordial non-Gaussianity, providing access to the cosmological collider.

In Chapter 3 we study the impact of the non-linear evolution of cosmic structure on constraints on primordial non-Gaussianity. Contrary to naive expectations, we reveal that even at very high redshift, where the matter density field is deemed to be highly linear, **non-Gaussian covariance** of the bispectrum triangle configurations significantly degrades sensitivity to local-type primordial non-Gaussianity. As an example, we demonstrate that failure to account for this additional covariance significantly overestimates the constraining power of a hydrogen intensity mapping survey (PUMA) at moderately high redshift.

The large non-Gaussian covariance of the bispectrum triangle configurations essentially states that the information that is present in these triangles in the linear den-

sity field, has been moved to higher-order correlation functions and smaller scales by non-linear evolution. In an attempt to recover the information, in Chapter 4 we study the use of artificial neural networks (ANNs) in **reconstructing the linear density field** from its late-time counterpart. Using a large suite of cosmological N-body simulations, we demonstrate that ANNs are indeed able to accurately reconstruct the linear density field. Furthermore, we show that the lowest-order statistics of the reconstructed density field contain significantly more information on cosmological parameters, including primordial non-Gaussianity, thus recovering valuable information present in the late-time density field.

In Chapter 5 we focus on the effect of the large-scale structure of the universe on observations of the CMB. As the light of the CMB travels throughout the universe, its path is deflected by the intervening matter distribution, an effect known as weak gravitational lensing. This effect slightly alters the statistics of the observed CMB from that of the emitted CMB, in particular obfuscating signatures of primordial non-Gaussianity. We develop a novel approach to **reconstruct the CMB lensing potential**, given lensed CMB observations, using state-of-the-art generative machine learning models, known as denoising diffusion models. Our method allows for a fully Bayesian treatment of the CMB lensing problem while being faster and more flexible than existing Bayesian methods.

Finally, in Chapter 6 we take a break from cosmology and study flat-space scattering amplitudes in effective field theories. We expose an intriguing relation between theories of **massive gravity and non-linear electrodynamics**. In particular, this involves theories that naturally appear in the scattering amplitude Double Copy web of theories, hinting at a possible role of massive gravity within this context.

Samenvatting

Enkele van de meest fundamentele open vragen in de natuurkunde hebben betrekking op wat er gebeurde tijdens de eerste momenten na de oerknal toen het universum ontstond. Een van de meest gevestigde theorieën is dat het heelal een periode van versnelde (exponentiële) uitdijing onderging gedurende een fractie van een seconde, waarbij de ruimte met meer dan twintig orden van grootte uitdijde. Deze periode wordt ook wel de **kosmische inflatie** genoemd. De meest overtuigende aanwijzingen dat zo'n uitdijing heeft plaatsgevonden, komen van de kosmische microgolf achtergrondstraling (Cosmic Microwave Background, CMB). Dit overgebleven licht, dat slechts 300.000 jaar na de oerknal werd uitgezonden, geeft ons een vroege momentopname van de staat van het vroege heelal. De exponentiële uitdijing helpt bij het verklaren van de waarneming dat twee causaal onverbonden gebieden in de CMB (bijv. twee tegenovergestelde richtingen aan de hemel) een bijna identieke temperatuur hebben. Deze schijnbare fijnafstemming van het universum kan in plaats daarvan worden verklaard als het waarneembare universum veel kleiner begon, zodat het geheel vroeg in causaal contact stond en zo een thermisch evenwicht kon bereiken. De CMB is echter niet perfect homogeen en isotroop en we hebben minuscule temperatuurschommelingen van 1 deel op 10000 gemeten, die Gaussisch verdeeld lijken te zijn. Opvallend genoeg biedt de kosmische inflatie theorie ook een mechanisme voor het genereren van deze kleine anisotropieën die we zien in de CMB, en die dienen als de zaden van structuurvorming en dus kunnen worden gezien als de begincondities van het universum, door middel van kwantumfluctuaties.

Inflatoire uitdijing kan op een groot aantal manieren fysisch worden gerealiseerd, waarvan vele verenigbaar zijn met onze nauwkeurige waarnemingen van de CMB tot nu toe. Om ons begrip van het vroege heelal te bevorderen, moeten we dus op zoek naar aanvullende manieren om verschillende inflatiescenario's te onder-

scheiden. Eén zo'n manier is via niet-Gaussische begincondities (**primordial non-Gaussianity**). Hoewel onze metingen van de CMB ons tot nu toe vertellen dat de begincondities Gaussisch lijken te zijn, voorspellen inflatiemodellen over het algemeen enige afwijking hiervan. Niet-Gaussische begincondities impliceren de aanwezigheid van correlatiefuncties van hogere orde dan de tweepuntscorrelatie die kenmerkend is voor een Gaussische verdeling. In tegenstelling tot de tweepuntscorrelatie (of het vermogensspectrum in de Fourier-ruimte), die alleen afhangt van de afstand tussen twee punten (of golflengte in de Fourier-ruimte), worden correlaties tussen meer dan twee punten weergegeven door veelhoeken. Opmerkelijk is dat verschillende inflatiemechanismen correlaties in verschillend gevormde veelhoeken teweegbrengen, waardoor deze als directe probe van inflatie gezien kunnen worden. Door ons bijvoorbeeld te richten op de laagste orde van niet-Gaussische correlaties, een driepuntsfunctie (of bispectrum in de Fourier-ruimte), kunnen we correlaties in verschillende vormen van driehoeken bestuderen, zoals gevisualiseerd in figuur 7.2.
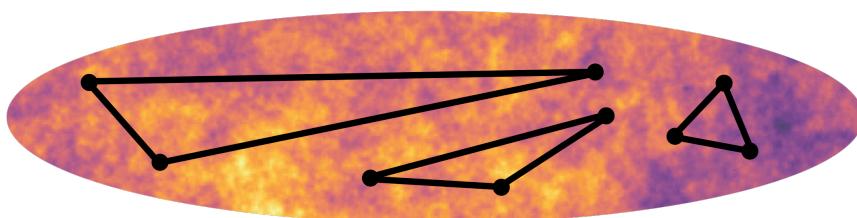

Figure 7.2: *De verdeling van primordiale dichtheidsfluctuaties, die dienen als de begincondities van ons universum, kan correlatiefuncties van hogere orde bevatten, zoals driepuntscorrelaties. Verschillende inflatiemechanismen induceren correlaties in verschillend gevormde driehoeken, zoals de samengeknepen, gelijkzijdige en platte vormen die hier worden gevisualiseerd.*

Het eenvoudigste theoretische model van kosmische inflatie vereist de aanwezigheid van een enkel scalair veld met een passend gevormde potentiële energiecurve, met minimale interactie met de zwaartekracht. De kwantumfluctuaties van dit scalaire veld vertalen zich in de primordiale dichtheidsveranderingen aan het einde van de inflatie. In dit scenario wordt voorspeld dat de afwijking van Gaussische begincondities klein is. Directe zelfinteracties van dit scalaire veld kunnen echter een aanzienlijk bispectrum genereren in gelijkzijdige driehoeksconfiguraties. Daarentegen zal de aanwezigheid van extra scalaire velden die bijdragen aan de inflatie-expansie waarschijnlijk een groot bispectrum in geknepen driehoeksconfiguraties veroorzaken. Het is dus duidelijk dat niet-Gaussische begincondities waardevolle inzichten kunnen geven in de dynamica van inflatie. Niet-Gaussische begincondities zijn niet

alleen een directe probe van het inflatiemechanisme, maar zijn ook gevoelig voor de interacties van het scalaire veld met andere velden die aanwezig zijn tijdens (maar niet bijdragen aan) inflatie, en waarvan de aanwezigheid een opvallende oscillatoire eigenschap induceert in de eerder genoemde geknepen driehoeken, waarvan de amplitude en frequentie gerelateerd is aan de massa en spin van het spectatorveld. Om deze reden kan inflatie worden beschouwd als een **kosmologische deeltjesversneller**, waarvan de uitkomst de begincondities van het universum heeft bepaald en waarvan de energieschaal ordes van grootte groter is dan die van welk terrestrisch deeltjesversnellerexperiment dan ook. Op deze manier bieden niet-Gaussische begincondities een uniek en krachtig inzicht in de fysica tijdens het inflatoire tijdperk.

Omdat we geen directe toegang hebben tot de begincondities, moeten we de eigenschappen afleiden uit het waarneembare universum dat uit deze begincondities voortkwam. Zoals eerder vermeld, levert de CMB een vrijwel onaangetaste blik op de primordiale dichtheidsfluctuaties en is daarom een onmisbare bron van informatie over het vroege heelal. De volgende generatie CMB-observaties heeft als doel om een nog hoger resolutiebeeld te krijgen van het vroege heelal, waarmee ons inzicht in niet-Gaussische begincondities zal worden verbeterd. Daarmee naderen deze onderzoeken echter de grens van wat de CMB ons kan vertellen over niet-Gaussische begincondities. We zijn dus gedwongen om onze aandacht te verleggen naar de verdeling van materie door het heelal, bekend als **de groteschaalstructuur van het universum**. Hoewel de driedimensionale verdeling van materie in principe vele malen meer informatie bevat dan de tweedimensionale CMB, wordt de relatie met de begincondities in de loop van de tijd niet-lineair door de gravitationele ineenstorting van de materie in gebonden structuren (zoals sterrenstelsels). Het extraheren van information over niet-Gaussische begincondities uit de materieverdeling gaat daarom gepaard met aanzienlijke moeilijkheden. Dit proefschrift presenteert verschillende vorderingen, uitdagingen en vooruitzichten in de jacht op niet-Gaussische begincondities in de groteschaalstructuur van het universum.

In hoofdstuk 2 onderzoeken we het gebruik van de 21-centimeter hyperfijne transitie van het neutrale waterstofgas dat het heelal vult tijdens de **kosmische donkere eeuwen**, als een probe van niet-Gaussische begincondities. We vinden dat deze vrijwel onaangetaste volger van primordiale fluctuaties, samen met zijn enorme driedimensionale volume, een ongeëvenaarde precisie oplevert om niet-Gaussische begincondities te detecteren of te beperken, hetgeen toegang geeft tot de kosmologische deeltjesversneller.

In hoofdstuk 3 bestuderen we de invloed van de niet-lineaire evolutie van kosmische structuren op het bepalen van niet-Gaussische begincondities. In tegen-

stelling tot naïeve verwachtingen, onthullen we dat zelfs op zeer hoge roodverschuiving, waar het materiedichtheidsveld geacht wordt zeer lineair te zijn, de **niet-Gaussische covariantie** van het bispectrum driehoeksconfiguraties de gevoeligheid voor niet-Gaussische begincondities van het lokale type aanzienlijk vermindert. Als voorbeeld laten we zien dat het niet in rekening brengen van deze extra covariantie de bepalende kracht van een waterstofspectrum intensiteitsmeting (PUMA) bij matig hoge roodverschuiving aanzienlijk overschat.

De grote niet-Gaussische covariantie van de driehoeksconfiguraties van het bispectrum geeft in wezen aan dat de informatie die in deze driehoeken aanwezig is in het lineaire dichtheidsveld, door niet-lineaire evolutie is verplaatst naar correlatiefuncties van hogere orde alsmede naar kleinere schalen. In een poging om de informatie terug te krijgen, bestuderen we in hoofdstuk 4 het gebruik van kunstmatige neurale netwerken (KNN's) om **het lineaire dichtheidsveld te reconstrueren** vanuit zijn latere niet-lineaire tegenhanger. Met behulp van een groot aantal kosmologische simulaties tonen we aan dat KNN's inderdaad in staat zijn om nauwkeurig het lineaire dichtheidsveld te reconstrueren. Verder laten we zien dat de laagste-orde statistiek van het gereconstrueerde dichtheidsveld significant meer informatie bevat over kosmologische parameters, inclusief niet-Gaussische begincondities, waardoor waardevolle informatie uit het niet-lineaire dichtheidsveld wordt teruggewonnen.

In hoofdstuk 5 richten we ons op het effect van de grootschalige structuur van het heelal op waarnemingen van de CMB. Als het licht van de CMB door het heelal reist, wordt zijn pad afgebogen door de tussenliggende materiedistributie, een effect dat bekend staat als zwakke gravitationele lensing. Dit effect verandert de statistiek van de waargenomen CMB enigszins ten opzichte van die van de uitgezonden CMB, waardoor met name tekenen van niet-Gaussische begincondities worden verdoezeld. We ontwikkelen **een nieuwe methode om het CMB lenspotentiaalveld te reconstrueren**, gegeven CMB-waarnemingen, met behulp van geavanceerde generatieve modellen, bekend als ontruizings diffusiemodellen. Onze methode maakt een volledig Bayesiaanse behandeling van het CMB lensing probleem mogelijk, terwijl het sneller en flexibeler is dan bestaande Bayesiaanse methoden.

Tenslotte nemen we in hoofdstuk 6 een pauze van de kosmologie en bestuderen we verstrooiingsamplitudes in effectieve veldentheorieën in vlakke ruimte-tijd. We leggen een intrigerende relatie bloot tussen theorieën van **massieve zwaartekracht en niet-lineaire elektrodynamica**. In het bijzonder gaat het om theorieën die van nature voorkomen in de verstrooiingsamplitude dubbel-kopie web van theorieën, wat wijst op een mogelijke rol van massieve zwaartekracht binnen deze context.

Acknowledgments

The completion of this thesis would not have been possible without the support of many people, whom I'd like to thank here.

Daan, being my most intensive supervisor and collaborator, I want to thank you for your time and dedication throughout these past four years. If it wasn't for you I would not be prepared for the horrors of inheriting a house and having to pay taxes over it for years! Jokes aside, you have taught me a great deal about cosmology and beyond. Hopefully, you have also been able to learn a thing or two from me. I greatly appreciated your habit of walking into my office three times a day, wondering "what's the story, morning glory!", promptly followed by you telling me what is actually the story. On top of that, I'm thankful for your (what appears to be) trust in me, by allowing me to work on whatever I wanted to work on, but also your efforts to put me in touch with many people in the field. Finally, my apologies for the times you had to endure my rants after hitting another bump in the road, when you surely had your own things to worry about. That said, I hope we will be able to keep working together in the future!

Diederik, the most casual certified genius, although our interactions became somewhat more rare this last year, I am very grateful for having had the chance to work with you. Even when after the first year I started spending more time on my computational research with Daan, you always remained excited and open to thinking about basically anything theoretical that we could come up with. Having perhaps been somewhat unlucky when it came to research topics and results, I am very happy that our names still appear together in at least one publication. Sometimes I wonder what else we could've discovered together, but then I realize I cannot derive the simplest equation. Finally, thank you for being such a positive and supportive person to everyone around you. It was a true pleasure sharing the corridor with you!

Léon, I think I owe you an apology for my clear inability to find the Kapteyn

Institute's entrance, or your office for that matter. Please allow me to blame it on the fact that a third direction of research would have been too much to handle for me. Nevertheless, you were always committed to your role as a supervisor and provided me with valuable feedback on various occasions, including of course on the thesis in front of you, no matter how busy your schedule. For this, I would like to thank you.

Thanks to Ruth Durrer, Michele Liguori, and Rien van de Weijgaert for being available and willing to assess this thesis.

Starting a PhD trajectory during the pandemic was not easy, but it was certainly made more enjoyable by my early collaborators. Tim, although your capacities and speed often made me insecure as your daily 'supervisor', this was largely offset by your modesty. You made crucial contributions to what would be our first scientific paper, for which I am very thankful. Tom, I enjoyed our hours spent scrutinising tedious derivations, looking at the ceiling for answers during hours of Zoom calls. After what seemed to take forever, I am glad that we eventually were able to put our names on a paper together with Diederik. I am glad that you managed to continue doing research, and look forward to reading about your undoubtedly meticulous investigations! On that same note, I would like to thank Jorik and Jelte for the fun interactions and collaboration during their Master's thesis research in our group. Thanks also to the other members of the Fundamentals of the Universe cohort; Martine, Dijs, Kevin, Giovanni, Jann and Bohdan.

During my PhD I was lucky to work with some great people. I would like to thank Matteo, Alba, Giorgio, Will, Paco, Adri, Ben and Oliver (Philcox), for many insightful discussion on physics and beyond during our collaborations. Additionally, I would like to thank Will and Paco for giving me the opportunity to visit the CCA in New York for several months. Adrian, Christoph, Florian, Gui, Oliver (Hahn) and many others that I am forgetting here, thank you for many interesting and useful discussions as well. Oliver, I am very grateful for the opportunity to continue my academic endeavours with your group in Vienna!

I want to thank the many members of the Van Swinderen Institute as well as the Zernike Institute theory group, for creating such a pleasant work environment. A special thanks to my office mates Martijn and Eric. Johannes, Ceyda, Evgenii, Alba, Gijsbert, Ivo, Sasha, thanks for the fun times spent outside of work as well.

Emma, Sjoerd, Mathijs, Gijs, Vincent, Rink, Floris, Levi, Joep, Tell 'em, Jonas, bedankt voor jullie steun en de welkome afleiding - middels zin dan wel onzin - door de jaren. Een extra woord van dank aan Gijs, voor het delen van academische ob-

sessies en de vele gesprekken die daaruit voortkwamen, en Vincent voor de nodige nuances en perspectieven op dit gebied, en het bijstaan tijdens de verdediging.

Meike, zonder jou als voorbeeld weet ik niet of ik überhaupt ooit het academisch pad was ingeslagen, en ik kan daarom geen passendere paranymf bedenken. Bedankt voor de onvoorwaardelijke steun, ook nadat ik weer eens weken niks van me laat horen. Papa en Mama, bedankt voor de vrijheid die ik altijd heb mogen ervaren om mijn eigen pad te bewandelen, vrij van druk en verwachtingen.

Vjesno, neizmerno hvala na ljubavi, pažnji i podršci tokom proteklih godina. Neopisivo sam zahvalan što si deo mog života i radujem se budućnosti sa tobom.

Curriculum Vitae

Thomas Stefan Flöss

4 December 1994	Born in Enschede, the Netherlands

Education

09/2020 - 08/2024	University of Groningen, the Netherlands Ph.D. research in Theoretical Cosmology Fundamentals of the Universe scholarship Supervised by prof. D. Roest, prof. L.V.E. Koopmans and dr. P.D. Meerburg
09/2016 - 08/2018	Utrecht University, the Netherlands Master Theoretical Physics Research project: *"Inflationary Consistency Conditions and Shift-Symmetric Cosmologies"* Supervised by dr. E. Pajer and dr. G.L. Goon
09/2013 - 08/2016	Utrecht University, the Netherlands Bachelor Physics and Astronomy Research project: *"Quantum Fluctuations and Magnon-Magnon Interactions in Antiferromagnets"* Supervised by prof. R.A. Duine and dr. S.A. Bender
09/2007 - 08/2013	Het Assink Lyceum, Haaksbergen, The Netherlands VWO, profiel Natuur en Techniek